Wide Area Information Servers: A Supercomputer on every Desk

Brewster Kahle
Thinking Machines Corporation

Wide Area Information Servers: A Supercomputer on every Desk

Brewster Kahle
Thinking Machines Corporation

What I really want...

- My personal information to be accessible
- Published information should find me
- Usable anywhere
- Others can use what I have learned (if I want them to)

What I really want...

- My personal information to be accessible
- Published information should find me
- Usable anywhere
- Others can use what I have learned (if I want them to)

What is it?

Electronic Publishing

(Or publishing over wires)

What is it?

Electronic Publishing

(Or publishing over wires)

Electronic Publishing

Professional searchers

$1/minute over obscure modems

//query (W5) inform?

600 databases on Dialog
~1 Terabyte
140Gbyte at DJ
80GB card catalog at RLG

Not understood

Electronic Publishing

Professional searchers

$1/minute over obscure modems

//query (W5)
inform?

600 databases
on Dialog
~1 Terabyte
140Gbyte at DJ
80GB card catalog
at RLG

Not understood

Telegraph> Telephone

Operators

Telephones on barb wire

Switching was manual

No white pages

Pay per minute

Telegraph> Telephone

Operators

Telephones on barb wire

Switching was manual

No white pages

Pay per minute

New Communications Technology Problems

	BOOKS	
Experts only	*Monks*	
Distribution is hard and expensive	*Vellum is calf skin*	
Different interfaces	*1000's of languages in Europe alone*	
Material is intractable	*Scrolls and manuscripts were about as random access as musical scores*	
Business model needed	*Centralized printing*	

New Communications Technology Problems

	BOOKS	
Experts only	*Monks*	
Distribution is hard and expensive	*Vellum is calf skin*	
Different interfaces	*1000's of languages in Europe alone*	
Material is intractable	*Scrolls and manuscripts were about as random access as musical scores*	
Business model needed	*Centralized printing*	

Navigation Techniques: Paper

- Alphabetical Listings (dictionary, Encyclopedia)
- Indices (back of the book and Readers Guide)
- Table of Contents (outlining)
- Citation index
- "Tree of Knowledge"
- Have you read any good books lately?

Navigation Techniques: Paper

- Alphabetical Listings (dictionary, Encyclopedia)
- Indices (back of the book and Readers Guide)
- Table of Contents (outlining)
- Citation index
- "Tree of Knowledge"
- Have you read any good books lately?

Navigation Techniques: Computers

- Hierarchical File Systems
- Unix "find" and "grep", Mac "find file"
- Boolean query systems (...within 5 words of...)
- Static Hypertext links (see also pointers)

Navigation Techniques: Computers

- Hierarchical File Systems
- Unix "find" and "grep", Mac "find file"
- Boolean query systems (...within 5 words of...)
- Static Hypertext links (see also pointers)

Navigation Techniques: WAIS

- English language questions and Relevance feedback
  - * Iterative retrieval
  - * Question-answer dialog
  - * Similar to the Newspapers front page the: "continued on page 5"
  - * Dynamic Hypertext Links
- 2 level search:
  - * Directory of servers (server like any other)
  - * Servers themselves
- Copy editors help select documents
  - * Easy to "publish" opinions on documents

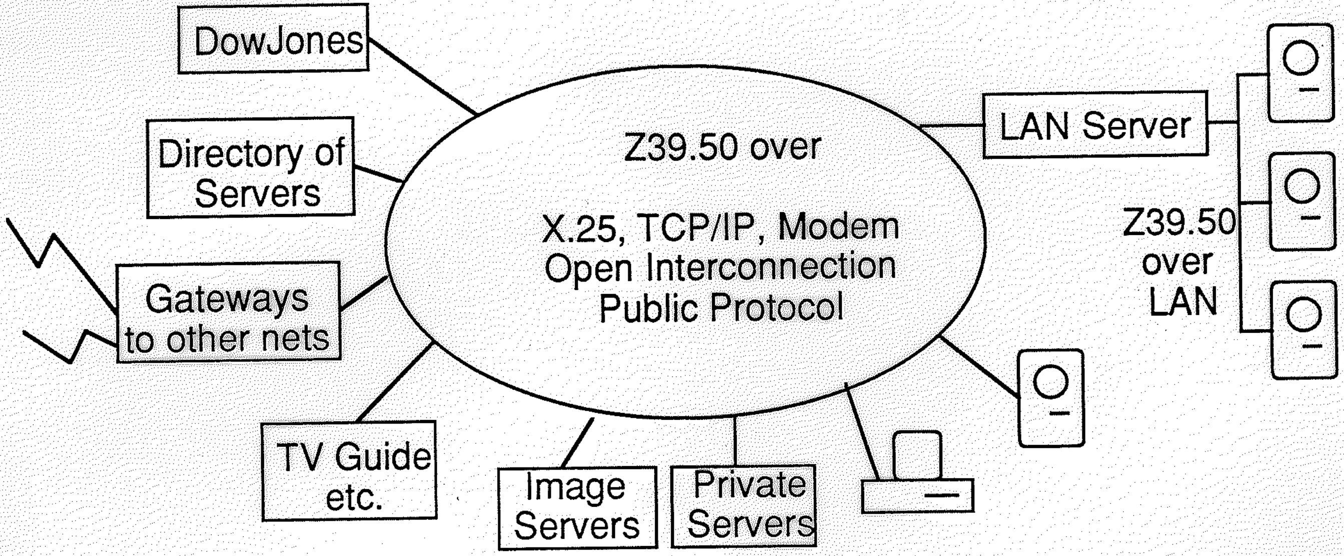
Navigation Techniques: WAIS

- English language questions and Relevance feedback
  * Iterative retrieval
  * Question-answer dialog
  * Similar to the Newspapers front page the: "continued on page 5"
  * Dynamic Hypertext Links

- 2 level search:
  * Directory of servers (server like any other)
  * Servers themselves

- Copy editors help select documents
  * Easy to "publish" opinions on documents

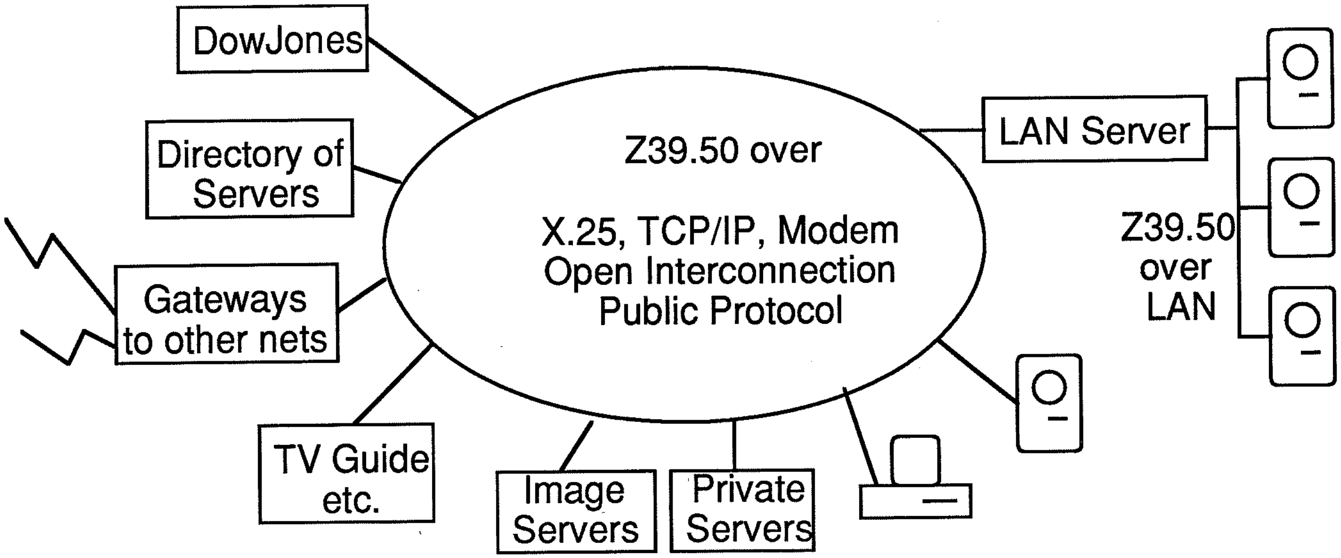
Wide Area Information Server Architecture



Users Needs:
- Selecting Servers
- Answering Questions
- Organizing Responses

Architecture Issues:
- Scalability
- Security
- Business model for servers
- Reliable Access

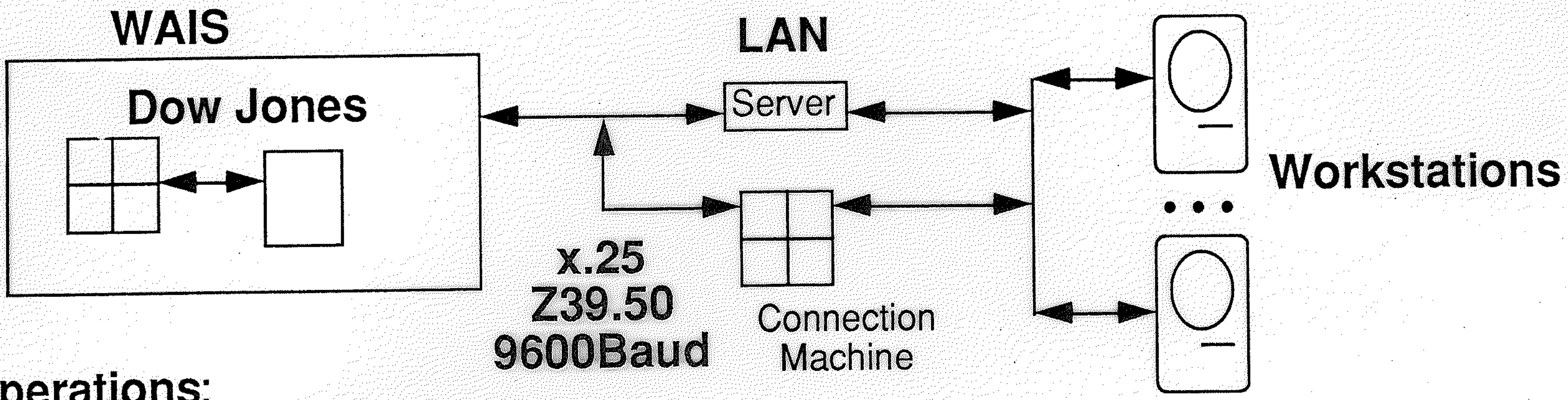
Wide Area Information Server Architecture



Users Needs:
- **Selecting Servers**
- **Answering Questions**
- **Organizing Responses**

Architecture Issues:
- **Scalability**
- **Security**
- **Business model for servers**
- **Reliable Access**

Demonstration System Structure



Operations:
- Archiving
- Queries
- Retrieval

IR Type:
- Broadcast
- Query by Example

Databases:
- Wall St Journal
- Barron's
- 400 Business Mags

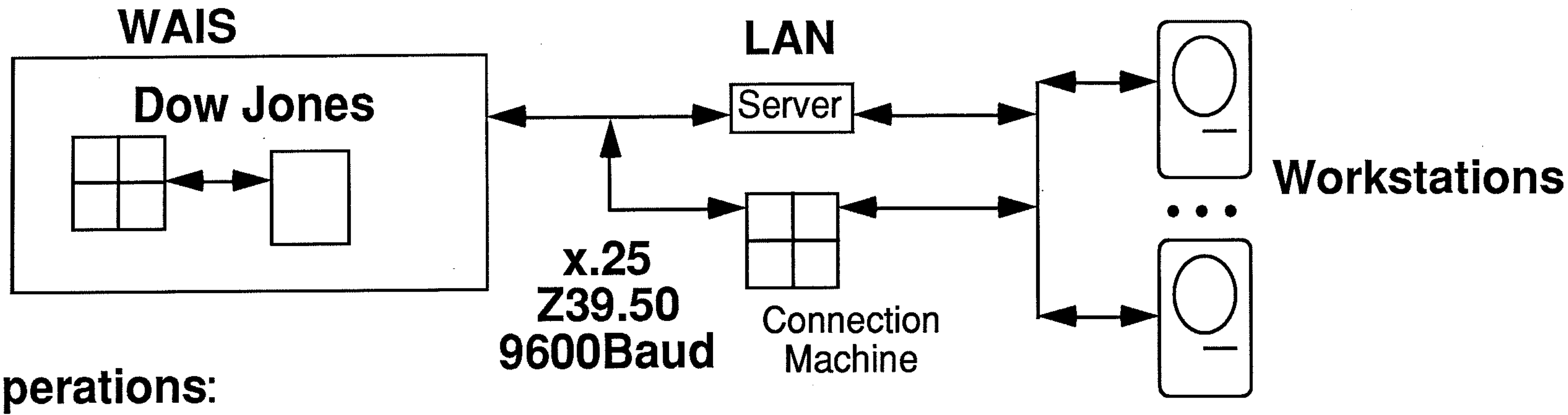
CM: Operations: Queries
IR Type:
enhanced relevance feedback
DBs: DowVision and
memo's, mail,
word processor files

Mac:
Operations:
- Human Int
- Retrieval
- Queries
- "Caching" Docs
- User Profiles

IR Type:
- Query by example

DBs:
- Personal Text
- Cached data

Demonstration System Structure

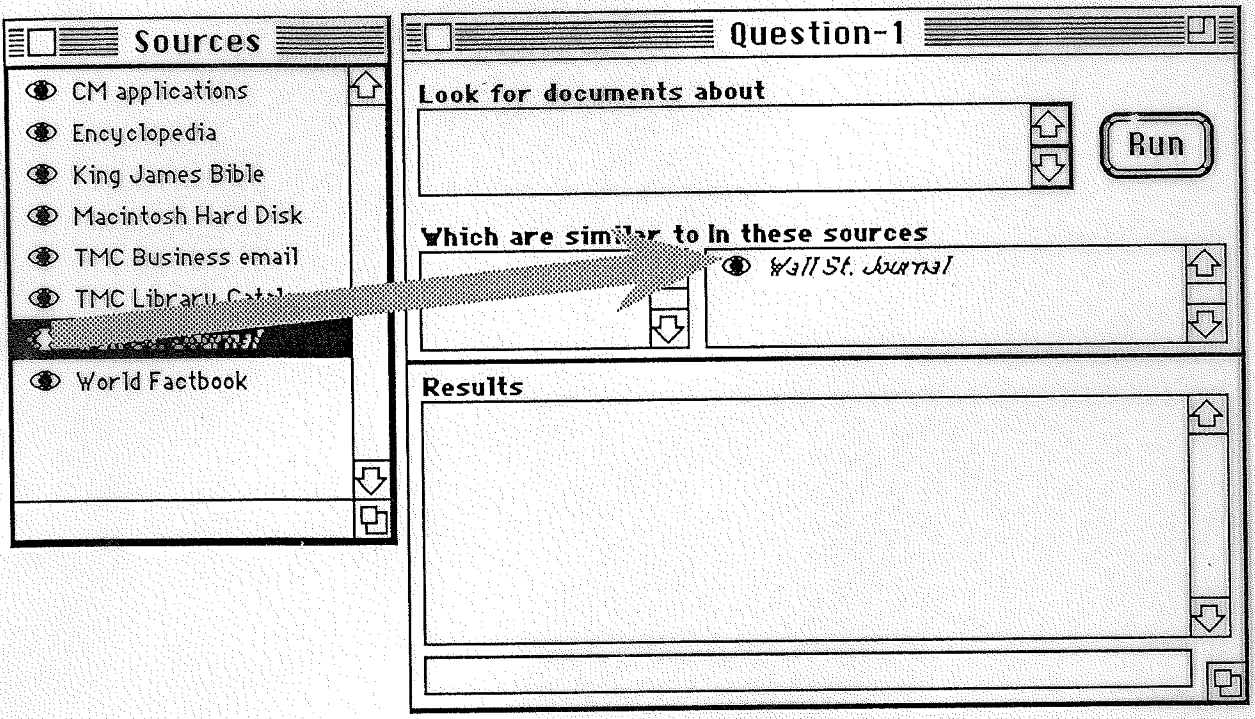


Operations:
Archiving
Queries
Retrieval
IR Type:
Broadcast
Query by Example
Databases:
Wall St Journal
Barron's
400 Business Mags

CM: **Operations**: Queries
IR Type:
enhanced relevance feedback
DBs: DowVision and
memo's, mail,
word processor files

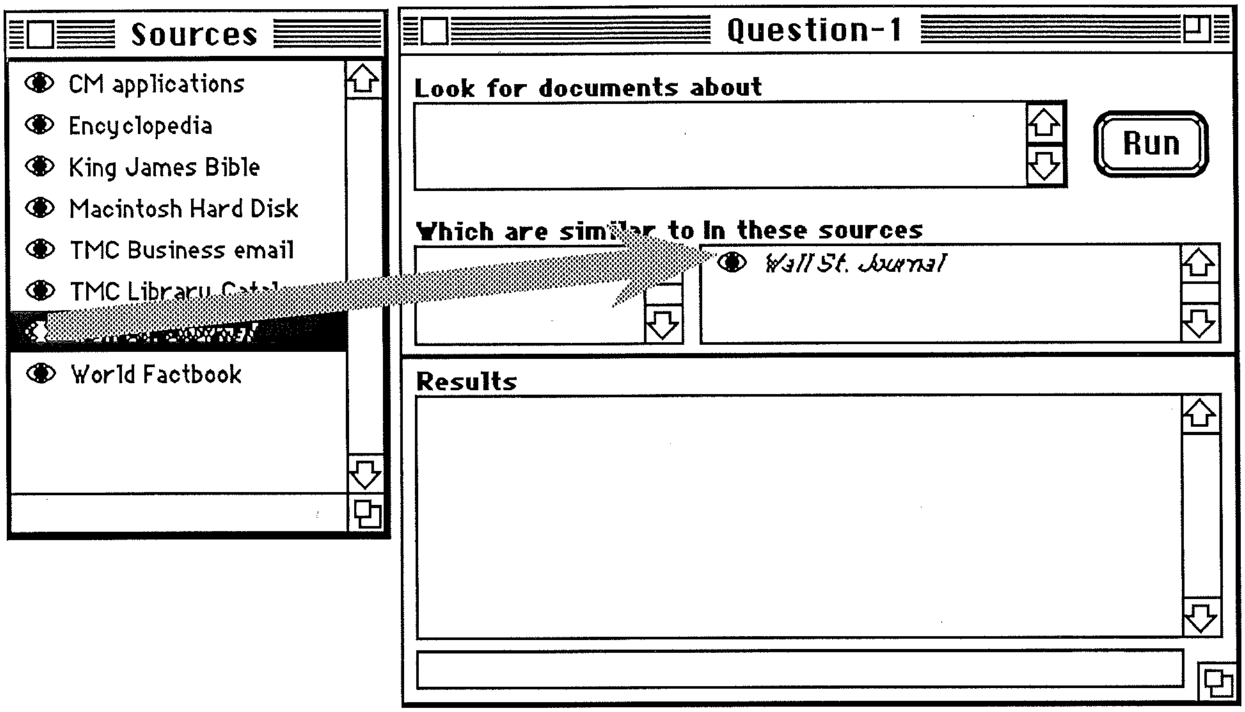
Mac:
Operations:
Human Int
Retrieval
Queries
"Caching" Docs
User Profiles
IR Type:
Query by example
DBs:
Personal Text
Cached data

WAIS Step 1



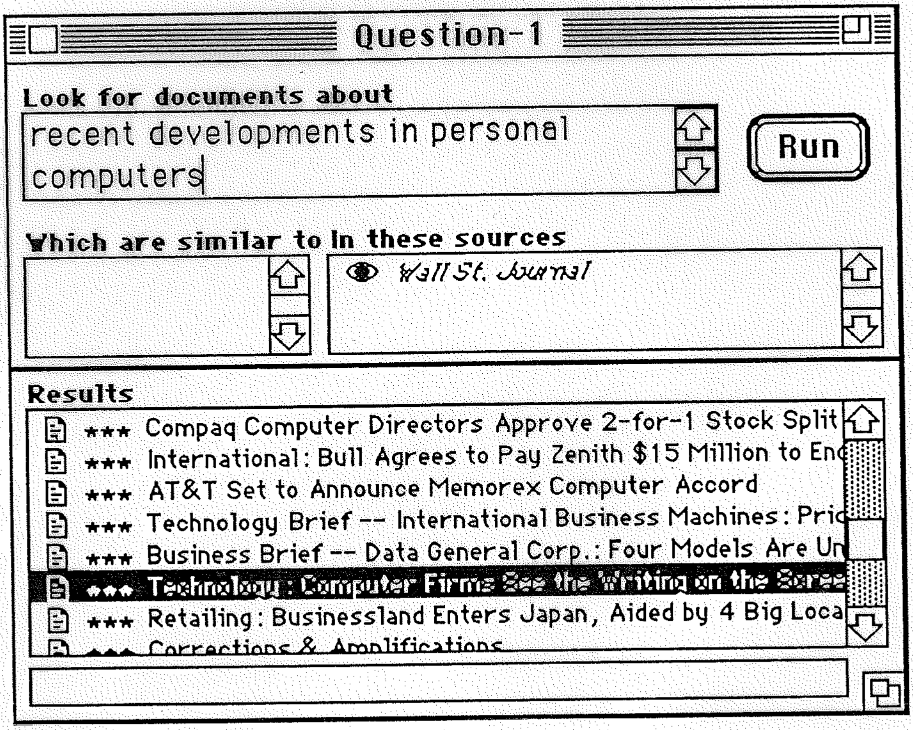
Step 1: Sources are dragged with the mouse into the Question Window. A question can contain multiple sources. When the question is run, it asks for information from each included source.

WAIS Step 1



Step 1: Sources are dragged with the mouse into the Question Window. A question can contain multiple sources. When the question is run, it asks for information from each included source.

WAIS Step 2



Question-1

Look for documents about

recent developments in personal computers

Which are similar to

In these sources

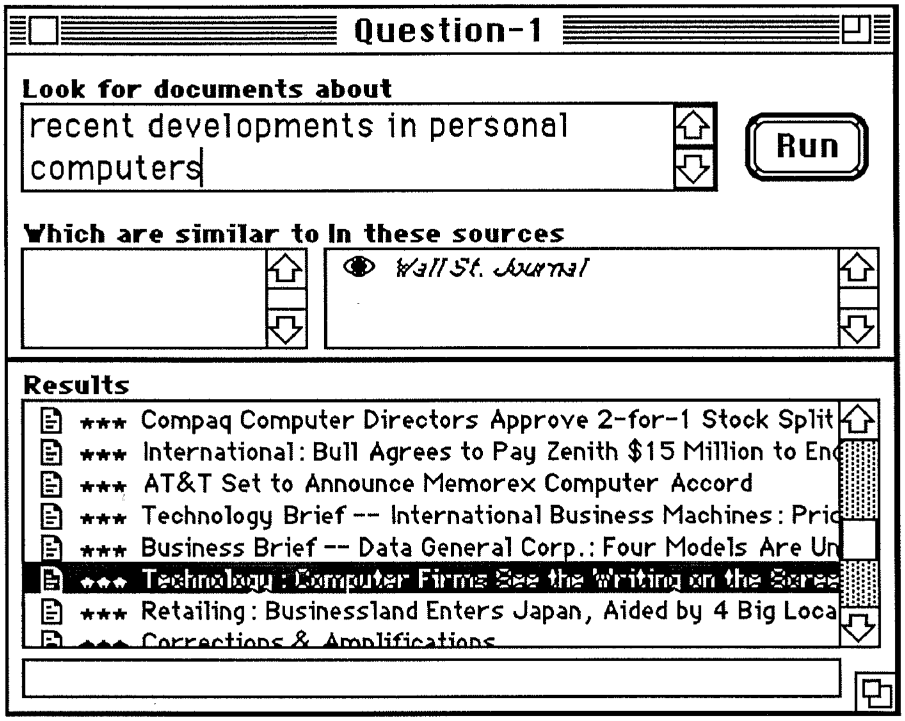
Wall St. Journal

Results

```
*** Compaq Computer Directors Approve 2-for-1 Stock Split
*** International: Bull Agrees to Pay Zenith $15 Million to En
*** AT&T Set to Announce Memorex Computer Accord
*** Technology Brief -- International Business Machines: Pri
*** Business Brief -- Data General Corp.: Four Models Are Ur
*** Technology: Computer Firms See the Writing on the Scree
*** Retailing: Businessland Enters Japan, Aided by 4 Big Loca
*** Corrections & Amplifications
```

Step 2: When a query is run, headlines of documents satisfying the query are displayed.

WAIS Step 2



Look for documents about

recent developments in personal computers

Which are similar to

In these sources

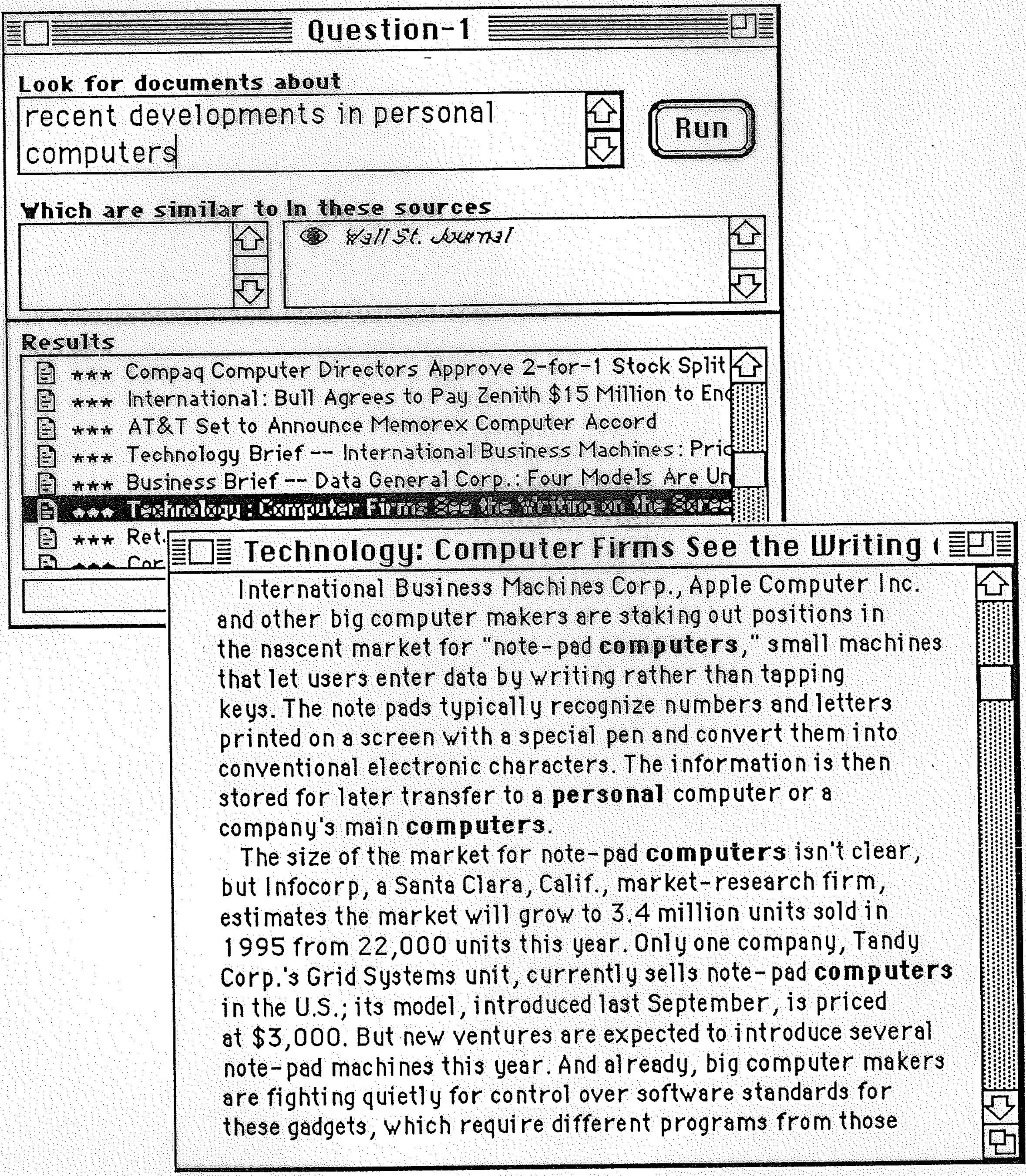
Wall St. Journal

Results

- *** Compaq Computer Directors Approve 2-for-1 Stock Split
- *** International: Bull Agrees to Pay Zenith $15 Million to Enc
- *** AT&T Set to Announce Memorex Computer Accord
- *** Technology Brief -- International Business Machines: Pric
- *** Business Brief -- Data General Corp.: Four Models Are Un
- *** Technology: Computer Firms See the Writing on the Scree
- *** Retailing: Businessland Enters Japan, Aided by 4 Big Loca
- *** Corrections & Amplifications

Step 2: When a query is run, headlines of documents satisfying the query are displayed.

WAIS Step 3



Look for documents about

recent developments in personal computers

Which are similar to

In these sources

Wall St. Journal

Results

- *** Compaq Computer Directors Approve 2-for-1 Stock Split
- *** International: Bull Agrees to Pay Zenith $15 Million to En
- *** AT&T Set to Announce Memorex Computer Accord
- *** Technology Brief -- International Business Machines: Pri
- *** Business Brief -- Data General Corp.: Four Models Are Un
- *** Technology: Computer Firms See the Writing on the Scree
- *** Ret
- *** Cor

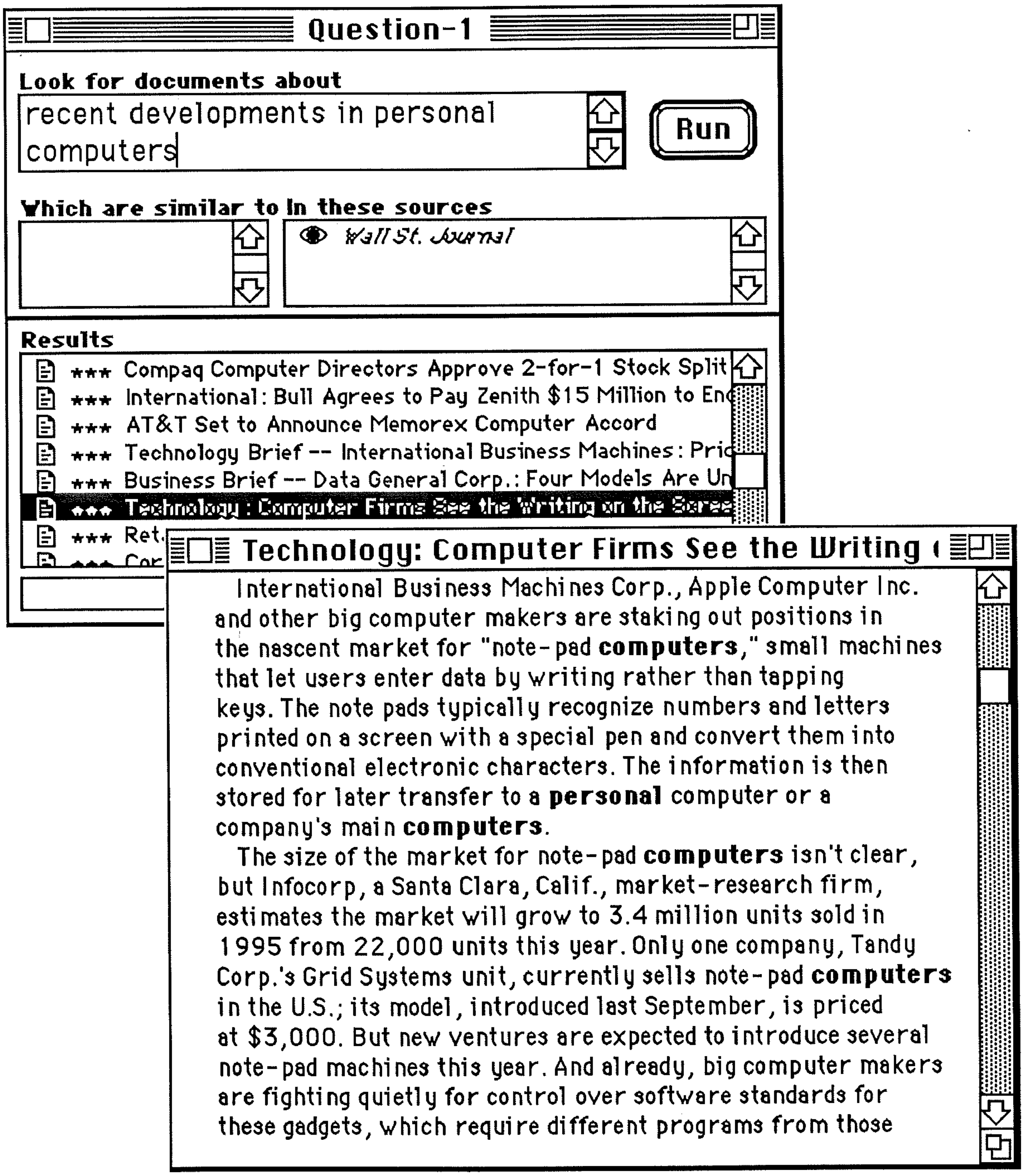
Technology: Computer Firms See the Writing

International Business Machines Corp., Apple Computer Inc. and other big computer makers are staking out positions in the nascent market for "note-pad **computers**," small machines that let users enter data by writing rather than tapping keys. The note pads typically recognize numbers and letters printed on a screen with a special pen and convert them into conventional electronic characters. The information is then stored for later transfer to a **personal** computer or a company's main **computers**.

The size of the market for note-pad **computers** isn't clear, but Infocorp, a Santa Clara, Calif., market-research firm, estimates the market will grow to 3.4 million units sold in 1995 from 22,000 units this year. Only one company, Tandy Corp.'s Grid Systems unit, currently sells note-pad **computers** in the U.S.; its model, introduced last September, is priced at $3,000. But new ventures are expected to introduce several note-pad machines this year. And already, big computer makers are fighting quietly for control over software standards for these gadgets, which require different programs from those

Step 3: With the mouse, the user clicks on any result document to retrieve it.

WAIS Step 3



Look for documents about

recent developments in personal computers

Which are similar to

In these sources

Wall St. Journal

- *** Compaq Computer Directors Approve 2-for-1 Stock Split
- *** International: Bull Agrees to Pay Zenith $15 Million to En
- *** AT&T Set to Announce Memorex Computer Accord
- *** Technology Brief -- International Business Machines: Pri
- *** Business Brief -- Data General Corp.: Four Models Are Un
- *** Technology: Computer Firms See the Writing on the Scree
- *** Ret
- *** Cor

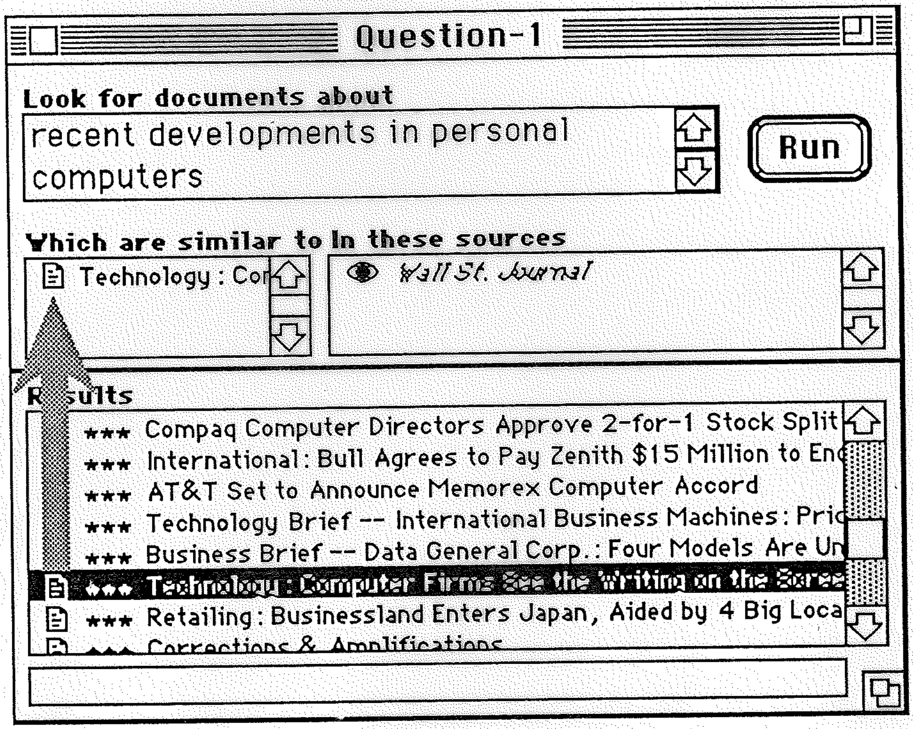
Technology: Computer Firms See the Writing

International Business Machines Corp., Apple Computer Inc. and other big computer makers are staking out positions in the nascent market for "note-pad **computers**," small machines that let users enter data by writing rather than tapping keys. The note pads typically recognize numbers and letters printed on a screen with a special pen and convert them into conventional electronic characters. The information is then stored for later transfer to a **personal** computer or a company's main **computers**.

The size of the market for note-pad **computers** isn't clear, but Infocorp, a Santa Clara, Calif., market-research firm, estimates the market will grow to 3.4 million units sold in 1995 from 22,000 units this year. Only one company, Tandy Corp.'s Grid Systems unit, currently sells note-pad **computers** in the U.S.; its model, introduced last September, is priced at $3,000. But new ventures are expected to introduce several note-pad machines this year. And already, big computer makers are fighting quietly for control over software standards for these gadgets, which require different programs from those

Step 3: With the mouse, the user clicks on any result document to retrieve it.

WAIS Step 4



Look for documents about: recent developments in personal computers

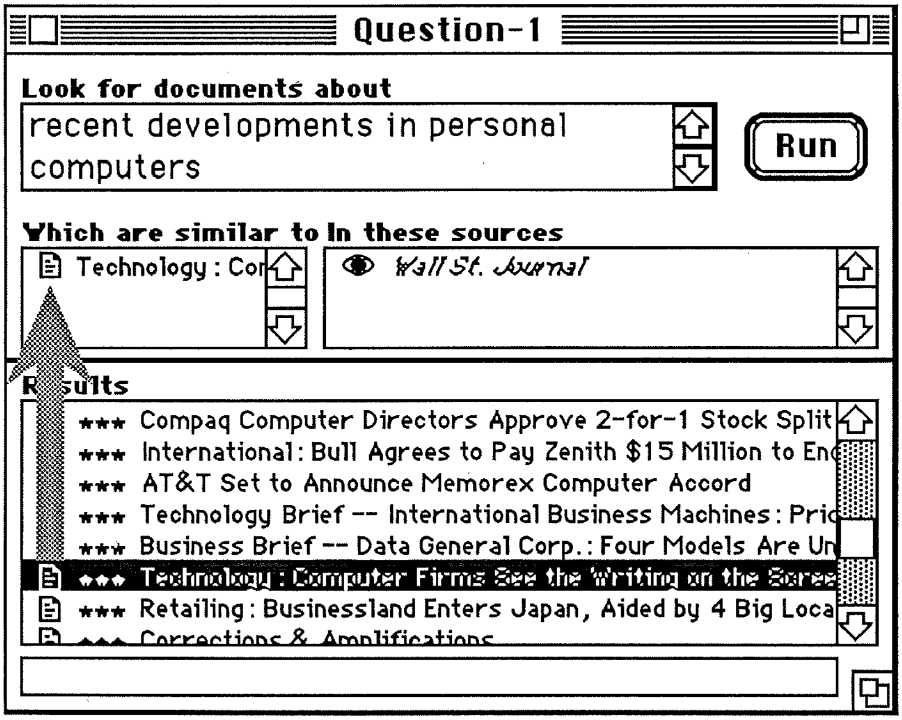
Which are similar to: Technology : Cor

In these sources: Wall St. Journal

- *** Compaq Computer Directors Approve 2-for-1 Stock Split
- *** International: Bull Agrees to Pay Zenith $15 Million to Enc
- *** AT&T Set to Announce Memorex Computer Accord
- *** Technology Brief -- International Business Machines: Pric
- *** Business Brief -- Data General Corp.: Four Models Are Un
- *** Technology : Computer Firms See the Writing on the Scree
- *** Retailing: Businessland Enters Japan, Aided by 4 Big Loca
- *** Corrections & Amplifications

Step 4: To refine the search, any one or more of the result documents can moved to the "Which are similar to:" box. When the search is run again, the results will be updated to include documents which are "similar" to the ones selected.

WAIS Step 4



Look for documents about

recent developments in personal computers

Which are similar to

Technology : Co

In these sources

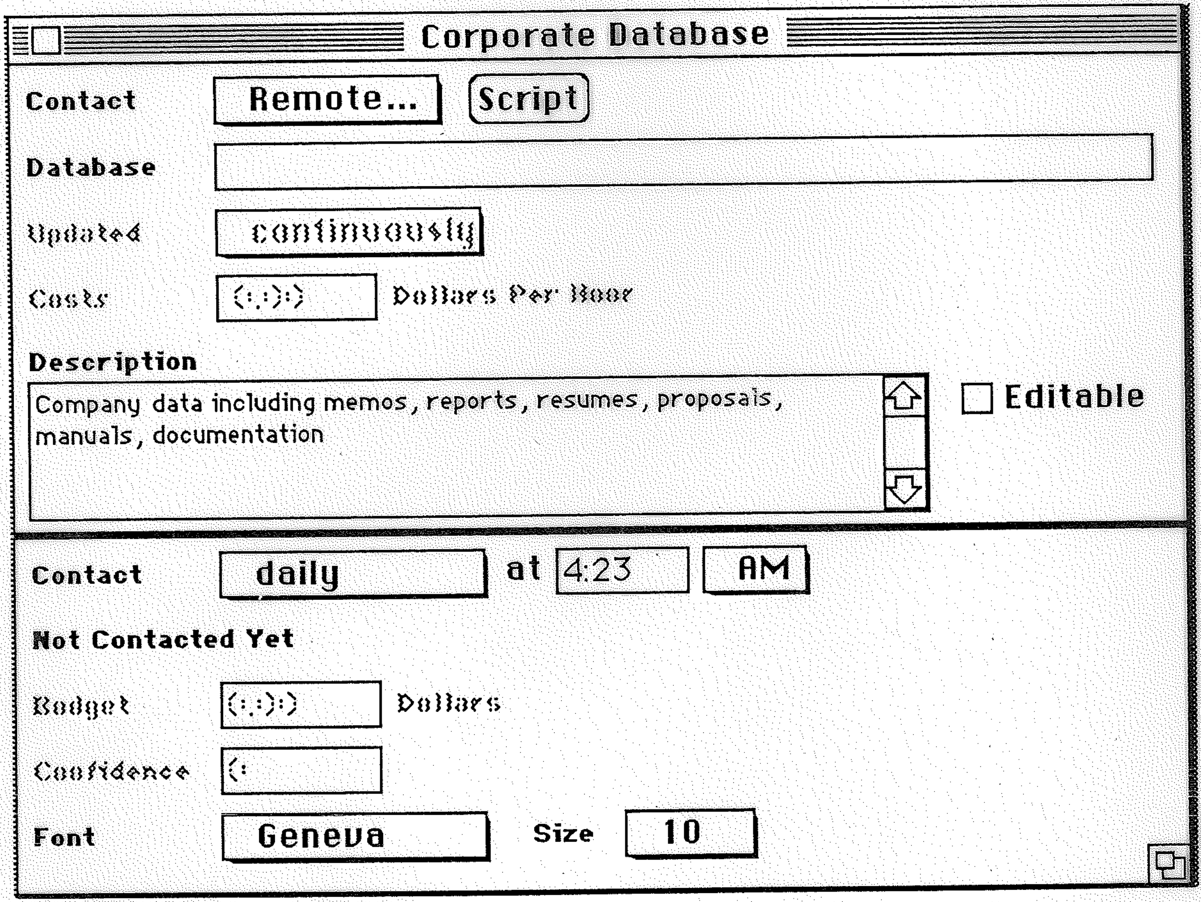
Wall St. Journal

Results

*** Compaq Computer Directors Approve 2-for-1 Stock Split
*** International: Bull Agrees to Pay Zenith $15 Million to En
*** AT&T Set to Announce Memorex Computer Accord
*** Technology Brief -- International Business Machines: Pric
*** Business Brief -- Data General Corp.: Four Models Are Un
*** Technology : Computer Firms See the Writing on the Scree
*** Retailing: Businessland Enters Japan, Aided by 4 Big Loca
*** Corrections & Amplifications

Step 4: To refine the search, any one or more of the result documents can moved to the "Which are similar to:" box. When the search is run again, the results will be updated to include documents which are "similar" to the ones selected.

Contacting Remote Sources of Information



Corporate Database

Contact: Remote... Script

Database:

Updated: continuously

Costs: (:.:):) Dollars Per Hour

Description: Company data including memos, reports, resumes, proposals, manuals, documentation

□ Editable

Contact: daily at 4:23 AM

Not Contacted Yet

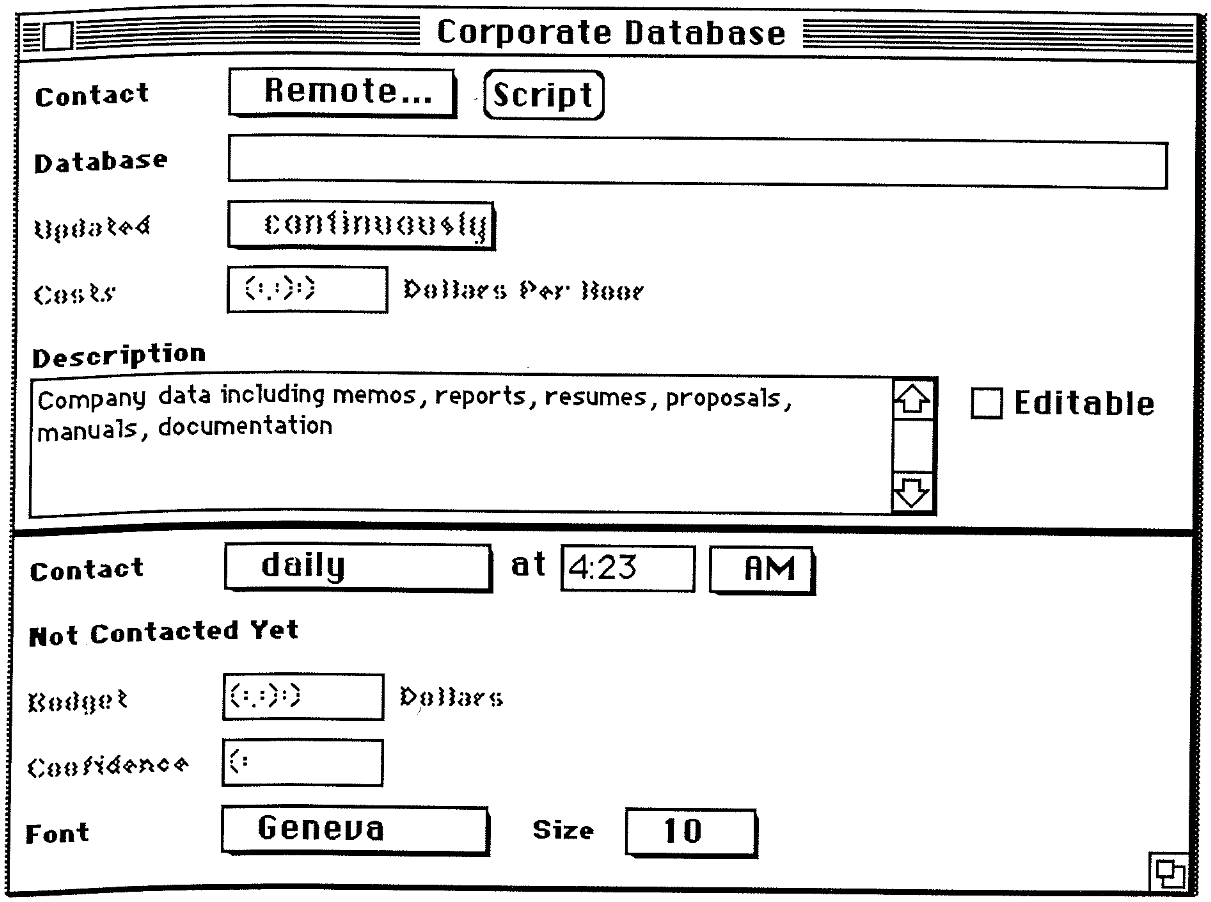
Budget: (:.:):) Dollars

Confidence: (:

Font: Geneva Size: 10

Figure 1: The Source description contains all the necessary information for contacting an information server.

Contacting Remote Sources of Information



Corporate Database

Contact: Remote... Script

Database:

Updated: continuously

Costs: 0.00 Dollars Per Hour

Description

Company data including memos, reports, resumes, proposals, manuals, documentation

□ Editable

Contact: daily at 4:23 AM

Not Contacted Yet

Budget: 0.00 Dollars

Confidence: 0

Font: Geneva Size: 10

Figure 1: The Source description contains all the necessary information for contacting an information server.

WAIS Clients

- Busy 24 hours a day finding information
- Ponder all indications of the preferences of its user
- Gossip with other clients about their discoveries
- Scours the world (within a budget) to find new sources

WAIS Clients

- Busy 24 hours a day finding information
- Ponder all indications of the preferences of its user
- Gossip with other clients about their discoveries
- Scours the world (within a budget) to find new sources

WAIS Protocol

- Based on Z39.50, bypass proprietary period
- Flexible
- Non Threatening for corporations

- Search: (words, doc_ids, databases) -> server
  returns list of: (headline, score, doc_id, types)'s
- Retrieval: (doc_id, type, start, end) -> server
  returns: bunch of bytes

- Doc_id: An ISBN for the Electronic Age
  ((orig_server, orig_database, orig_local_id)
  (dist_server, dist_database, dist_local_id)
- Server Description:
  (:ip-address, :database-name, :cost, :description)

WAIS Protocol

- Based on Z39.50, bypass proprietary period
- Flexible
- Non Threatening for corporations

- Search: (words, doc_ids, databases) -> server
  returns list of: (headline, score, doc_id, types)'s
- Retrieval: (doc_id, type, start, end) -> server
  returns: bunch of bytes

- Doc_id: An ISBN for the Electronic Age
  ((orig_server, orig_database, orig_local_id)
  (dist_server, dist_database, dist_local_id)
- Server Description:
  (:ip-address, :database-name, :cost, :description)

Connection Machine Server

- 1-25GBytes (and getting bigger)
- Supports thousands of users
- Automatic Indexing
- Uses words and phrases in question to find appropriate documents
- First turn-key massively parallel application

Connection Machine Server

- 1-25GBytes (and getting bigger)
- Supports thousands of users
- Automatic Indexing
- Uses words and phrases in question to find appropriate documents
- First turn-key massively parallel application

TMC Internet Release

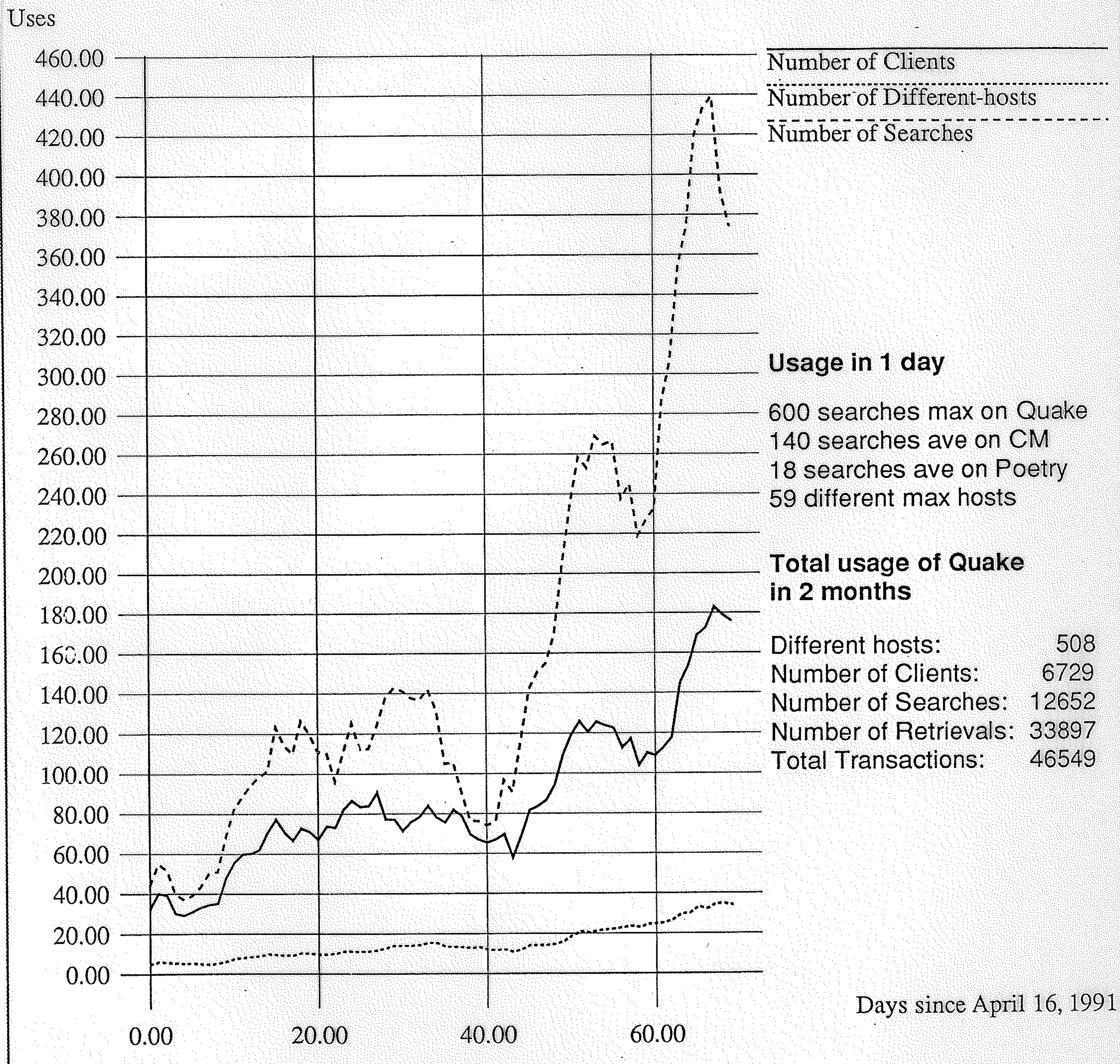
- CM product for TCP/IP (complete server)
- Example User interfaces for free (no support)
  Macintosh, Gnu Emacs, Xwindows
- Example unix server software to create servers
- Directory of Servers on the internet at least through '91
- 25 Servers now: Weather Maps, patents, Government programs, Risks-digest, usenet recipies, Lewis Carrol,...
- Anonymous FTP Think.com:/public/wais/*
  Mailing list: wais-discussion-request@think.com

TMC Internet Release

- CM product for TCP/IP (complete server)
- Example User interfaces for free (no support)
  Macintosh, Gnu Emacs, Xwindows
- Example unix server software to create servers
- Directory of Servers on the internet at least through '91
- 25 Servers now: Weather Maps, patents, Government programs, Risks-digest, usenet recipies, Lewis Carrol,...
- Anonymous FTP Think.com:/public/wais/*
  Mailing list: wais-discussion-request@think.com

WAIS

WAIS Daily Usages on Quake.Think.Com



Usage in 1 day

600 searches max on Quake
140 searches ave on CM
18 searches ave on Poetry
59 different max hosts

Total usage of Quake in 2 months

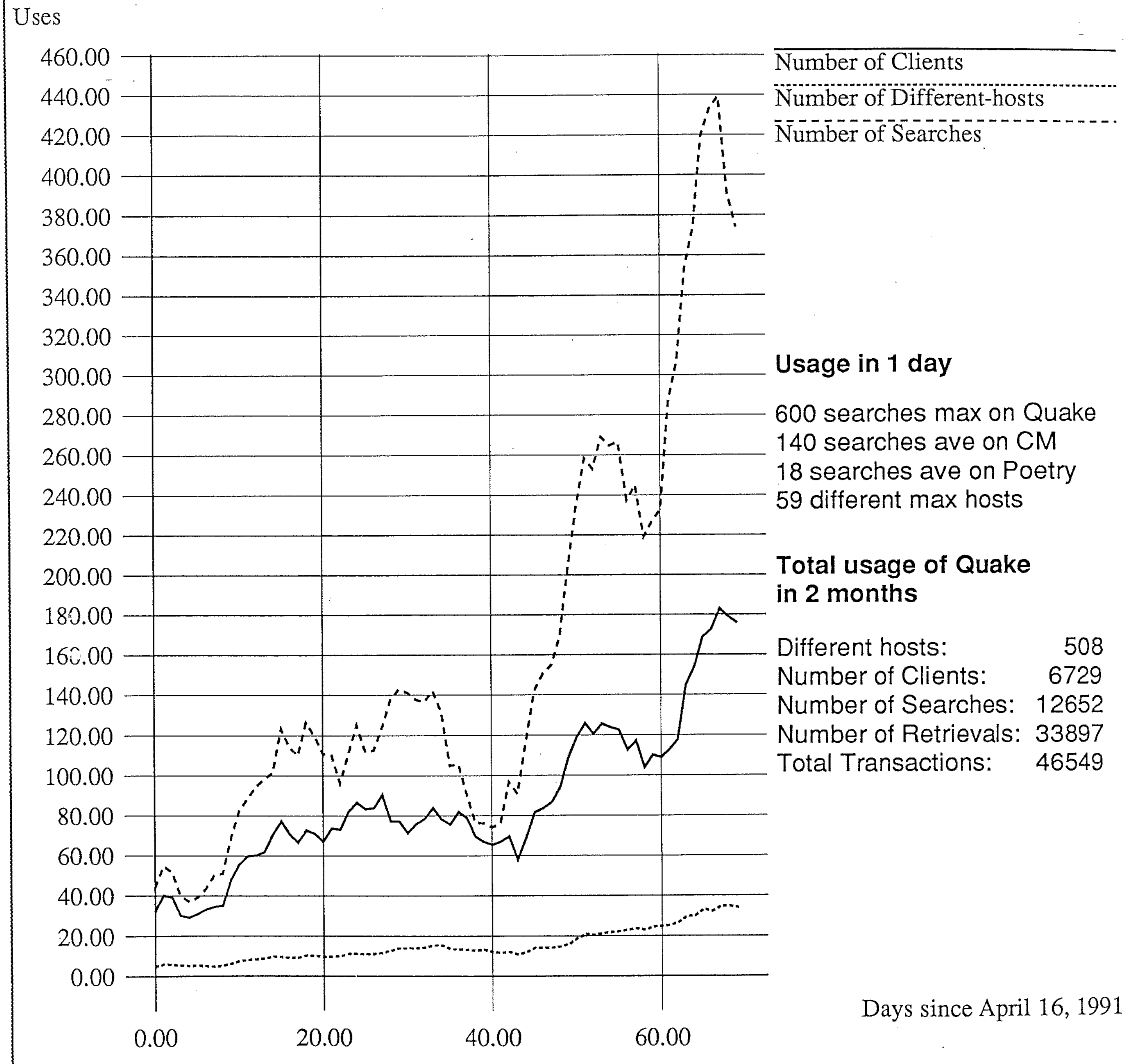
Different hosts:	508
Number of Clients:	6729
Number of Searches:	12652
Number of Retrievals:	33897
Total Transactions:	46549

Countries Using WAIS:
Austria, Canada, Denmark, Finland, France, Germany, Holland, Italy, Mexico, Norway, Sweden, Switzerland, USA

Thinking Machines Corporation

WAIS

WAIS Daily Usages on Quake.Think.Com



Usage in 1 day

600 searches max on Quake
140 searches ave on CM
18 searches ave on Poetry
59 different max hosts

Total usage of Quake in 2 months

Different hosts:	508
Number of Clients:	6729
Number of Searches:	12652
Number of Retrievals:	33897
Total Transactions:	46549

Countries Using WAIS:
Austria, Canada, Denmark, Finland, France, Germany, Holland, Italy, Mexico, Norway, Sweden, Switzerland, USA

Thinking Machines Corporation

WAIS Servers

Top level server of servers (maintained by Thinking Machines):
directory-of-servers.src

Connection Machine documentation (servers on Connection Machine):
CM-fortran-manual.src CM-paris-manual.src CM-star-lisp-docs.src
CM-tech-summary.src CMFS-documentation.src CM-applications.src

MIT algorithms book adendum (servers at MIT):
MIT-algorithms-bug.src MIT-algorithms-exercise.src
MIT-algorithms-suggest.src

Internet directories etc (servers at NSF and Thinking Machines)
internet-documents.src internet-drafts.src internet-resource-guide.src
internet-rfcs.src

PD programs for mainframes (server in georgia)
cosmic-abstracts.src cosmic-programs.src US-Gov-Programs.src

Picture servers:
sample-pictures.src weather.src

Mail archive servers (various places):
jik-usenet.src sun-spots.src risks-digest.src
homebrew.src info-mac.src

Server in Olso Norway:

UiO_Publications.src	;;Research interests of professors

Library catalogs (various places):
tmc-library.src online-libraries.src

Servers on WAIS:
wais-discussion-archives.src wais-docs.src

Misc.

Molecular-biology.src	;;genetics abstracts
NIH-Guide.src	;;guide to RFP's
bible.src	;;King James Bible
usenet-cookbook.src	;;Cookbook
jargon.src	;;Hacker's Dictionary
world-factbook.src	;;CIA descriptions of countries
poetry.src	;;Shakespeare, Yeats, Sawyer, etc
patent-sampler.src	;;20Mbytes of patents (full text)
rkba.src	;;Right to keep and bear arms documents
sample-books.src	;;A few books such as Lewis Carroll's etc
wall-street-journal-sample.src	;;Couple of months from 1989 WSJ

WAIS Servers

Top level server of servers (maintained by Thinking Machines):
directory-of-servers.src

Connection Machine documentation (servers on Connection Machine):
CM-fortran-manual.src CM-paris-manual.src CM-star-lisp-docs.src
CM-tech-summary.src CMFS-documentation.src CM-applications.src

MIT algorithms book adendum (servers at MIT):
MIT-algorithms-bug.src MIT-algorithms-exercise.src
MIT-algorithms-suggest.src

Internet directories etc (servers at NSF and Thinking Machines)
internet-documents.src internet-drafts.src internet-resource-guide.src
internet-rfcs.src

PD programs for mainframes (server in georgia)
cosmic-abstracts.src cosmic-programs.src US-Gov-Programs.src

Picture servers:
sample-pictures.src weather.src

Mail archive servers (various places):
jik-usenet.src sun-spots.src risks-digest.src
homebrew.src info-mac.src

Server in Olso Norway:
UiO_Publications.src ;;Research interests of professors

Library catalogs (various places):
tmc-library.src online-libraries.src

Servers on WAIS:
wais-discussion-archives.src wais-docs.src

Misc.

Molecular-biology.src	;;genetics abstracts
NIH-Guide.src	;;guide to RFP's
bible.src	;;King James Bible
usenet-cookbook.src	;;Cookbook
jargon.src	;;Hacker's Dictionary
world-factbook.src	;;CIA descriptions of countries
poetry.src	;;Shakespeare, Yeats, Sawyer, etc
patent-sampler.src	;;20Mbytes of patents (full text)
rkba.src	;;Right to keep and bear arms documents
sample-books.src	;;A few books such as Lewis Carroll's etc
wall-street-journal-sample.src	;;Couple of months from 1989 WSJ

Conclusion

- Electronic Publishing can fill niches now
- Companies are positioning themselves now (workstations, server, and info providers)
- Thinking Machines is the "Engine of the Information Industry"

Conclusion

- Electronic Publishing can fill niches now
- Companies are positioning themselves now (workstations, server, and info providers)
- Thinking Machines is the "Engine of the Information Industry"

Wide Area Information Servers: A Supercomputer on every Desk

Brewster Kahle
Thinking Machines Corporation

What I really want...

- My personal information to be accessible
- Published information should find me
- Usable anywhere
- Others can use what I have learned (if I want them to)

What is it?

Electronic Publishing

(Or publishing over wires)

New Communications Technology Problems

	BOOKS	Telegraph> Telephone	Electronic Publishing
Experts only	*Monks*	*Operators*	*Professional searchers*
Distribution is hard and expensive	*Vellum is calf skin*	*Telephones on barb wire*	*$1/minute over obscure modems*
Different interfaces	*1000's of languages in Europe alone*	*Switching was manual*	*//query (W5) inform?*
Material is intractable	*Scrolls and manuscripts were about as random access as musical scores*	*No white pages*	*600 databases on Dialog* *~1 Terabyte* *140Gbyte at DJ* *80GB card catalog at RLG*
Business model needed	*Centralized printing*	*Pay per minute*	*Not understood*

Navigation Techniques: Paper

- Alphabetical Listings (dictionary, Encyclopedia)
- Indices (back of the book and Readers Guide)
- Table of Contents (outlining)
- Citation index
- "Tree of Knowledge"
- Have you read any good books lately?

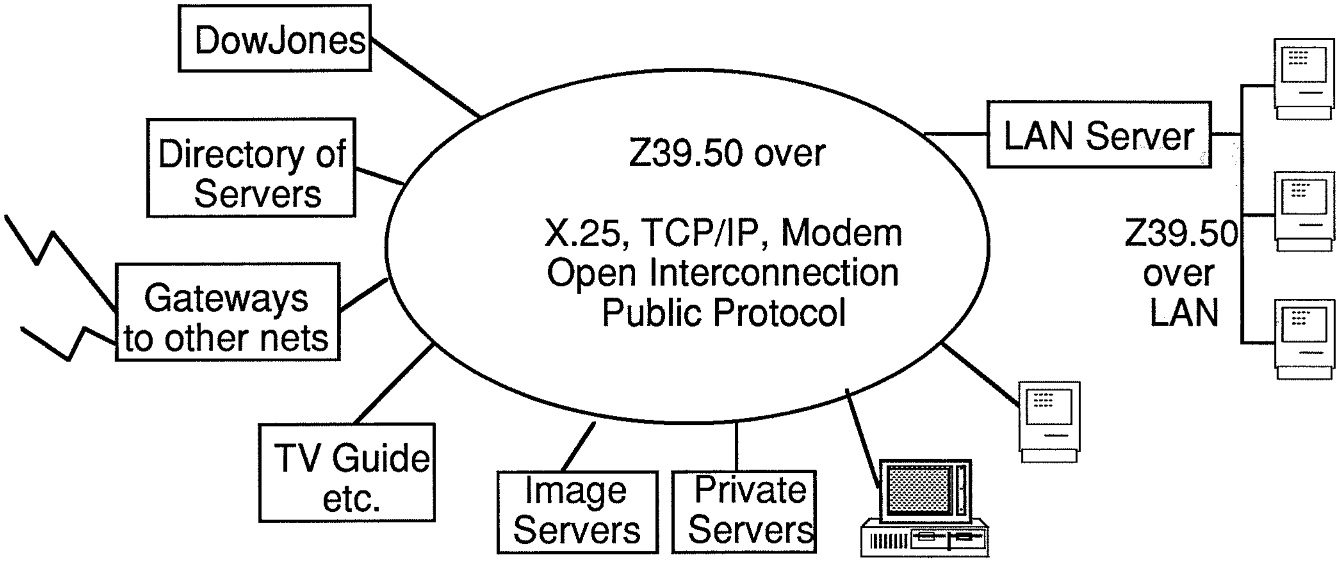
Navigation Techniques: Computers

- Hierarchical File Systems
- Unix "find" and "grep", Mac "find file"
- Gopher, Magellan, ON Location
- Boolean query systems (...within 5 words of...)
- Static Hypertext links (see also pointers)

Navigation Techniques: WAIS

- English language questions and Relevance feedback
  - Question-answer dialog
  - Similar to Newspapers: "More on page 5"
  - Dynamic Hypertext Links
- 2 level search:
  - Directory of servers (server like any other)
  - Servers themselves

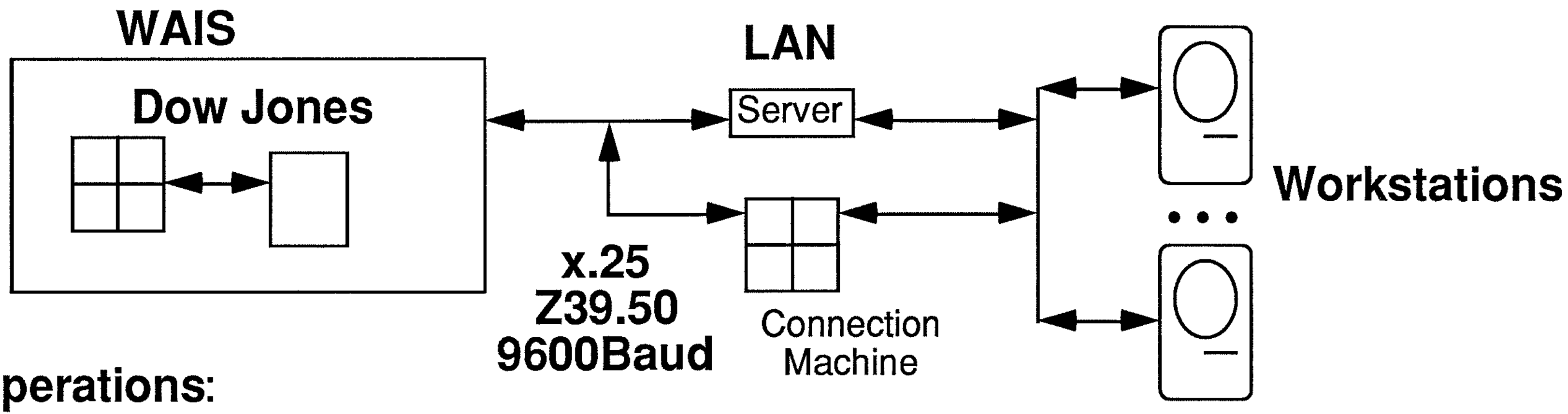
Wide Area Information Server Architecture



Users Needs:
- **Selecting Servers**
- **Answering Questions**
- **Organizing Responses**

Architecture Issues:
- **Scalability**
- **Security**
- **Business model for servers**
- **Reliable Access**

Peat Marwick System Structure



WAIS

Dow Jones

LAN

Server

Workstations

x.25
Z39.50
9600Baud

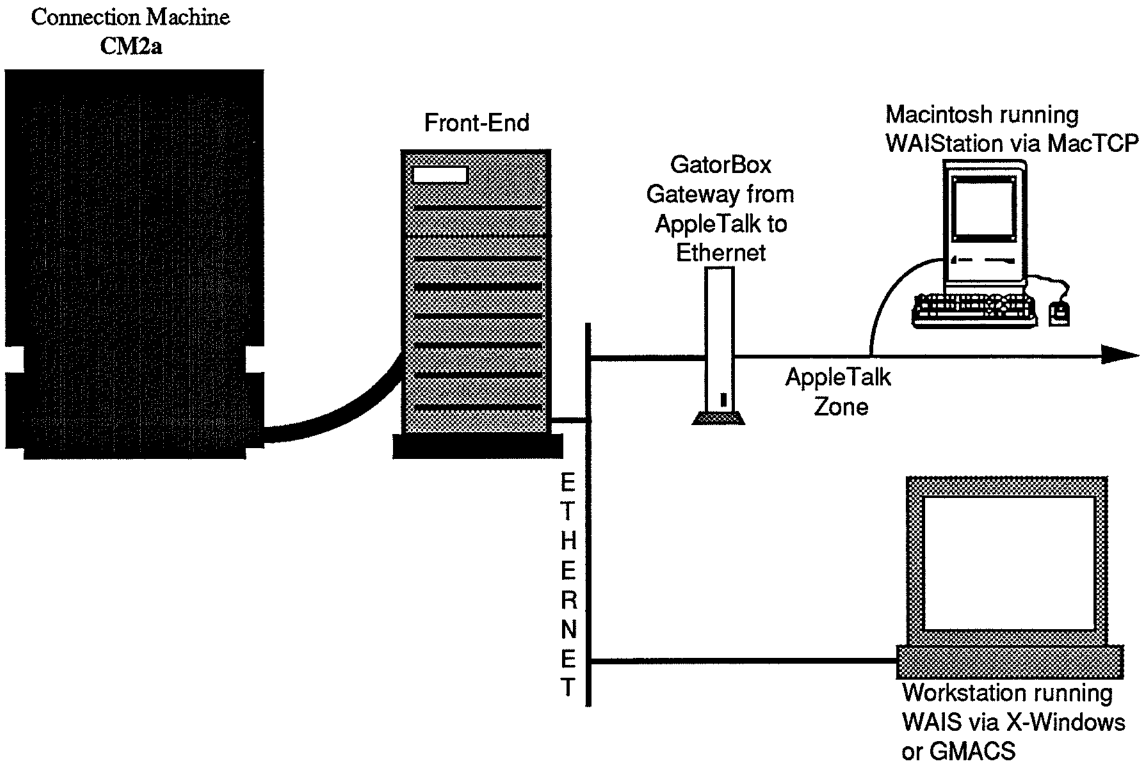
Connection
Machine

Operations:
Archiving
Queries
Retrieval
IR Type:
Broadcast
Query by Example
Databases:
Wall St Journal
Barron's
400 Business Mags

CM: **Operations**: Queries
IR Type:
enhanced relevance feedback
DBs: DowVision and
memo's, mail,
word processor files

Mac:
Operations:
Human Int
Retrieval
Queries
"Caching" Docs
User Profiles
IR Type:
Query by example
DBs:
Personal Text
Cached data

WAIS Hardware Components



WAIS Clients

- Busy 24 hours a day finding information
- Ponder all indications of the preferences of its user
- Gossip with other clients about their discoveries
- Scours the world (within a budget) to find new sources
- Current implementations on PC, Macintosh,
  X Windows, NeXT, dumb terminal (dial-up)

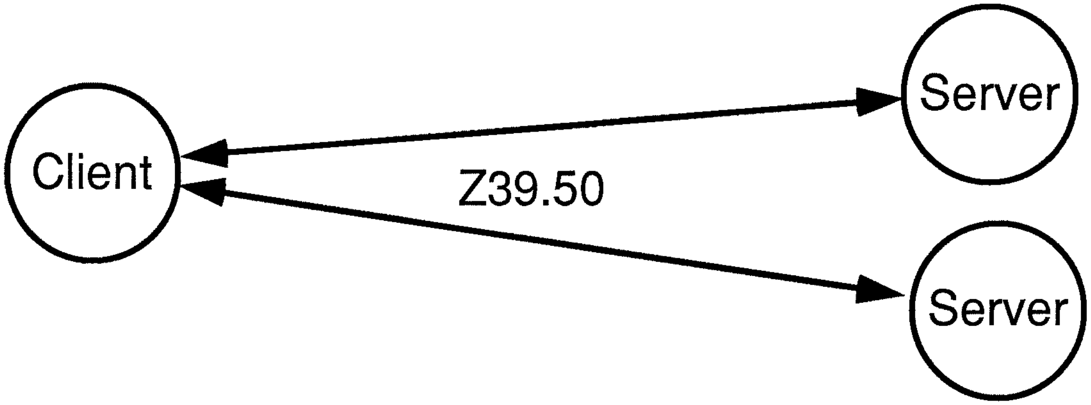
WAIS Protocol

- Based on NISO Z39.50 international standard
- Flexible — separates clients from servers
- Search: (words, doc_ids, databases) returns list of: (headline, score, doc_id, types)
- Retrieval: (doc_id, type, start, end) returns: data of specified type
- Doc_id: An ISBN for the Electronic Age
- Server Description Structure for the Directory of Servers

How Standard Protocol can Provide Security

- Users do not login to server, but search only through application layer protocol (Z39.50)
- Server controls access to data
- Network layers below application, or application layer handles authentication, encryption, billing

The WAIS Protocol *is* WAIS



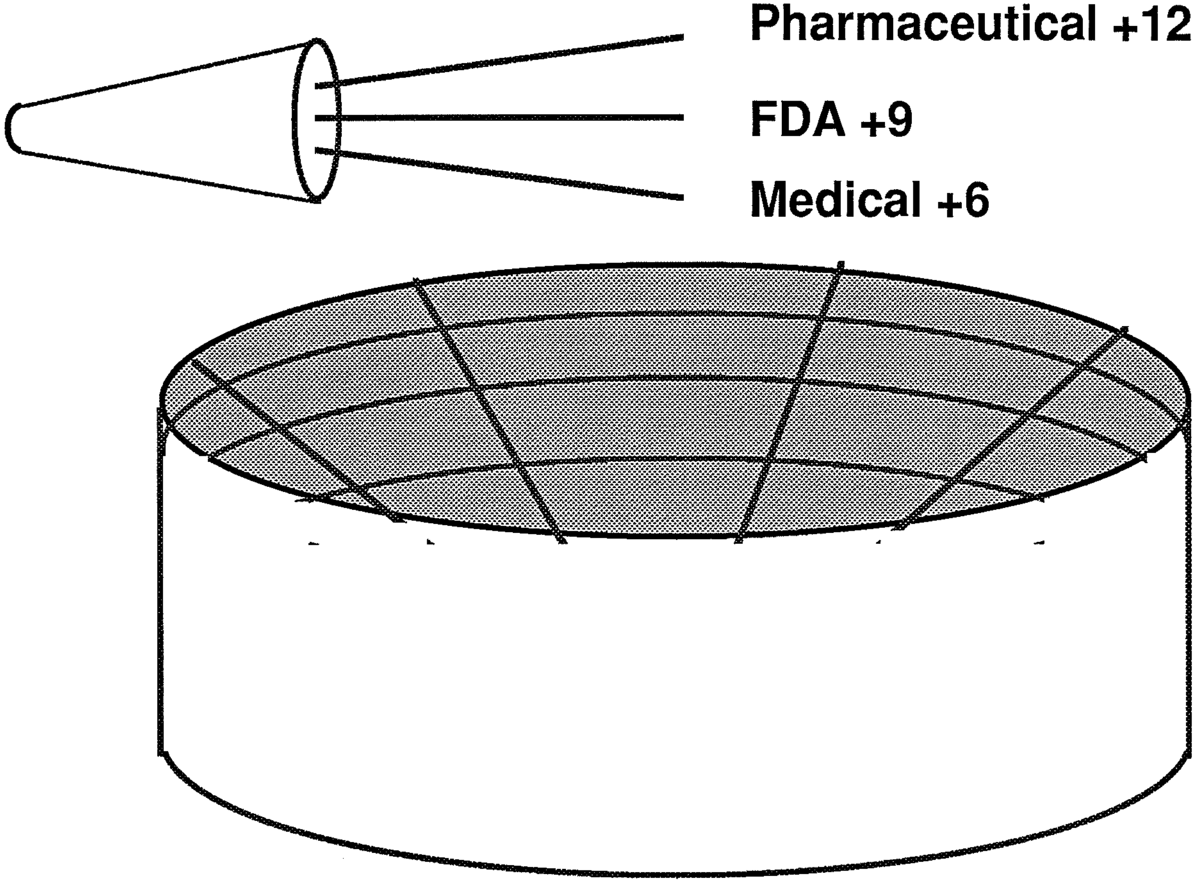
- Supports any search syntax
- Supports sophisticated clients — puts intelligence in the user's hands
- Clients can run on any platform
- Multiple servers in a single search
- Retrieve any kind of data: text, graphics, video,…

Connection Machine Server

- 1-100GBytes (and getting bigger)
- Supports thousands of users
- Automatic Indexing
- Uses words and phrases in question to find appropriate documents with relevance feedback, weighted term
- Supports Boolean Queries
- Cost effective hardware alternative to mainframes

Data Parallelism:

Searching all the documents at once



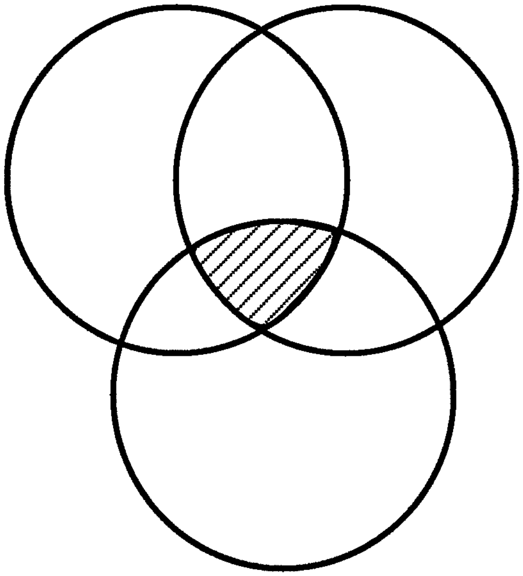
Pharmaceutical +12

FDA +9

Medical +6

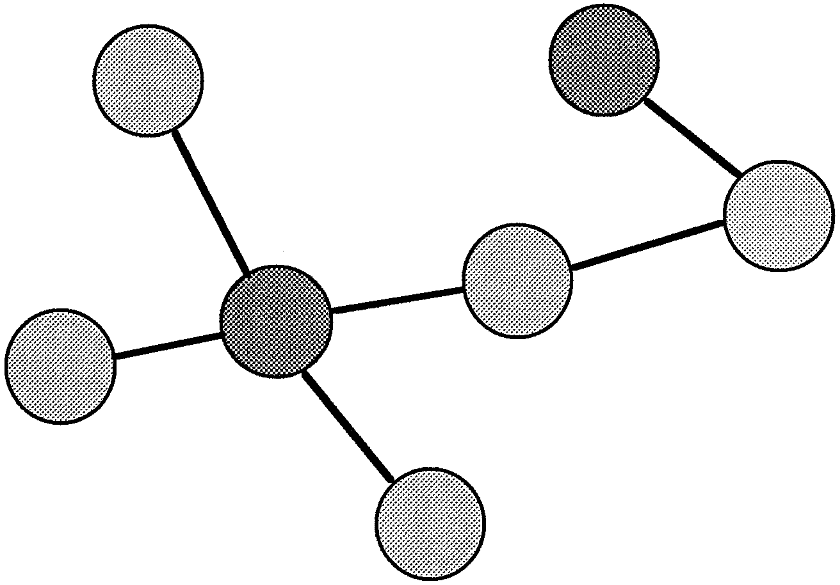
Stadium

Boolean Search



Retrieve documents containing specific combinations of words

Conceptual Search



Explore a set of documents containing related concepts

Boolean Query

Hard to Use:

Complex Syntax

```
(Japanese OR Japan) AND
(building OR buildings OR (Real AND Estate) AND
(Manhattan OR (New AND York)
```

Poor Results:

The wrong information

No ranking of results

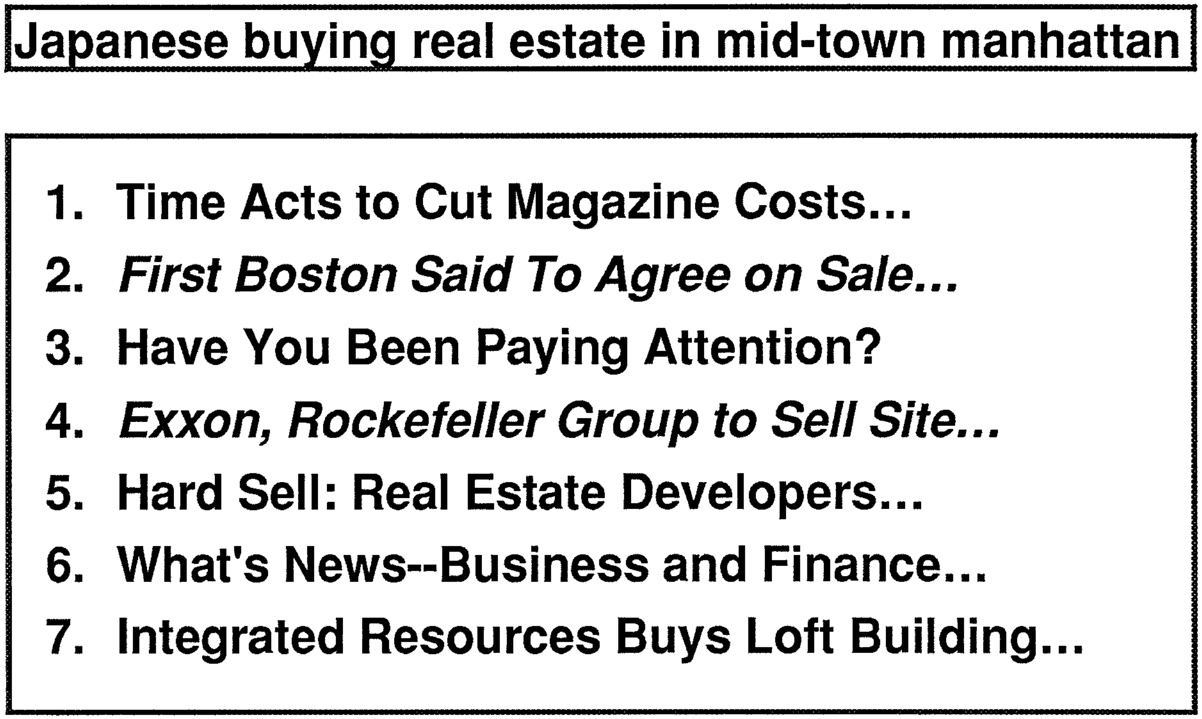
Have you been paying attention?...
Freer Finance: U.S. Regulators Move...
REAL ESTATE: California Initiatives...
First Boston Said To Agree on Sale Of...
Exxon, Rockefeller Group to Sell Site...
What's News--Business and Finance

Conceptual Search: Phase 1

Easy to Use:
No Syntax

Options:

What do you want to follow up?



Japanese buying real estate in mid-town manhattan

1. Time Acts to Cut Magazine Costs...
2. *First Boston Said To Agree on Sale...*
3. Have You Been Paying Attention?
4. *Exxon, Rockefeller Group to Sell Site...*
5. Hard Sell: Real Estate Developers...
6. What's News--Business and Finance...
7. Integrated Resources Buys Loft Building...

Conceptual Search: Phase 2

Relevance Feedback:

I like these; show me more

First Boston Said To Agree on Sale...

Exxon, Rockefeller Group to Sell Site...

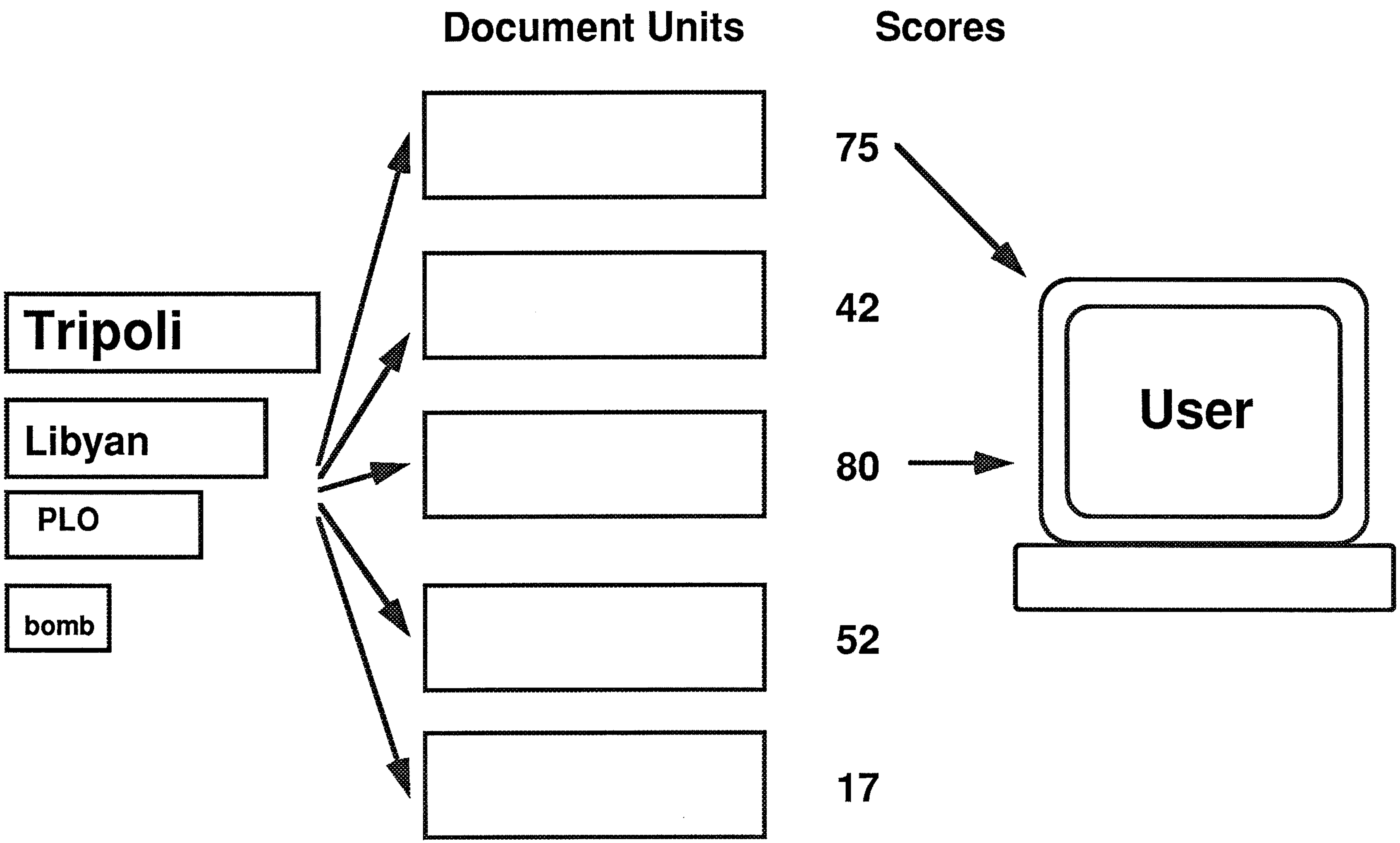
Improved results:

Articles on related topics are found

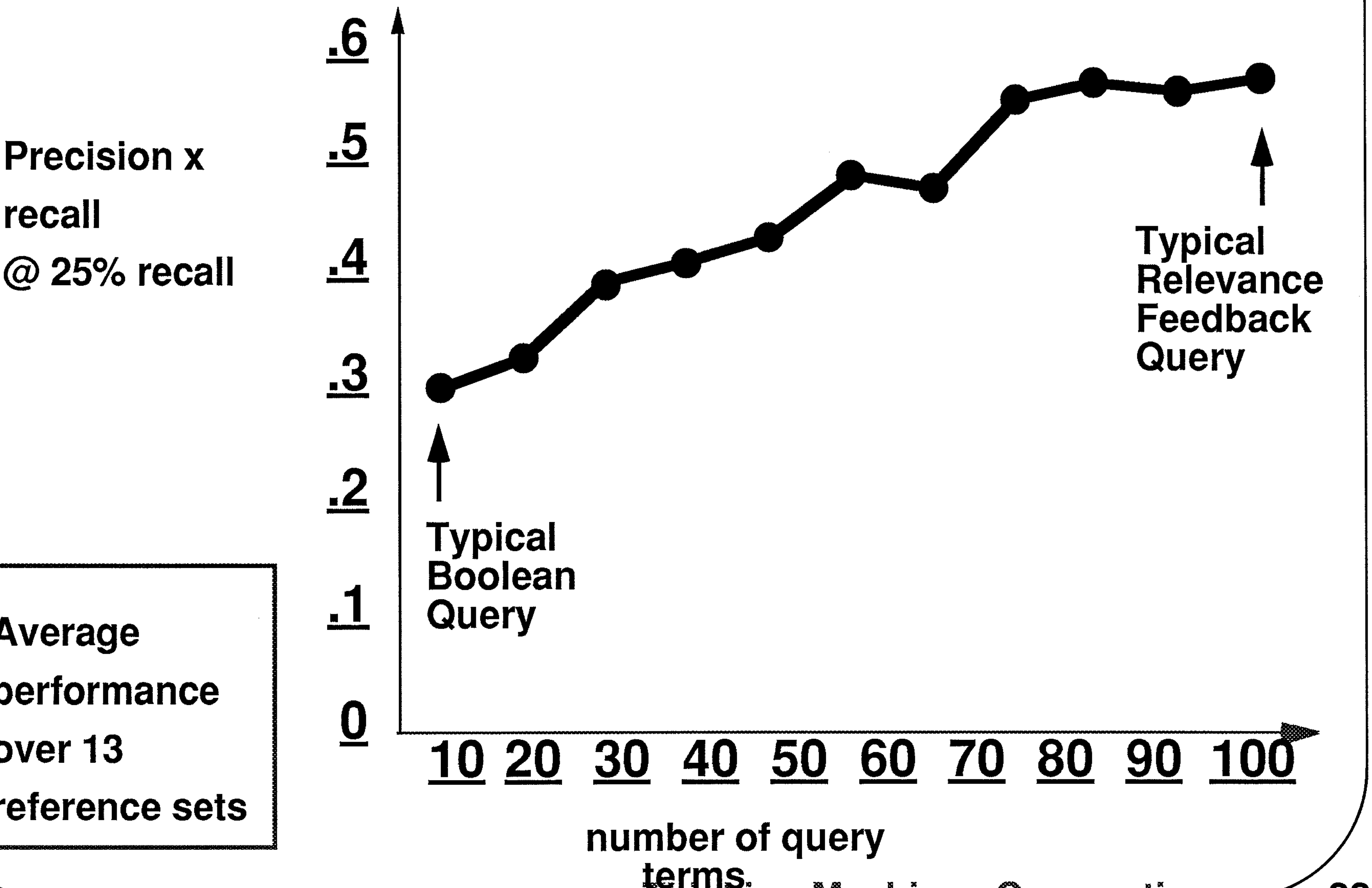
Results are ranked

1. Bids for Exxon Building in New York...
2. Time Acts to Cut Magazine Costs...
3. Hard Sell: Real Estate Developers...
4. Time Inc. Sells Its 45% Interest...
5. Citicorp Unit Moves to Foreclose on...
6. Litigious Landlords: Legal Maneuvers

Query Broadcast To Database on Connection Machine System



Results Improve with Query Size



Average performance over 13 reference sets

Document Retrieval Performance

- Current algorithm limits:
  ~2 GB with 512 MB CM-2
  ~8 GB with 2 GB CM-2
  ~25 GB with 8 GB CM-2

- High recall
- High precision

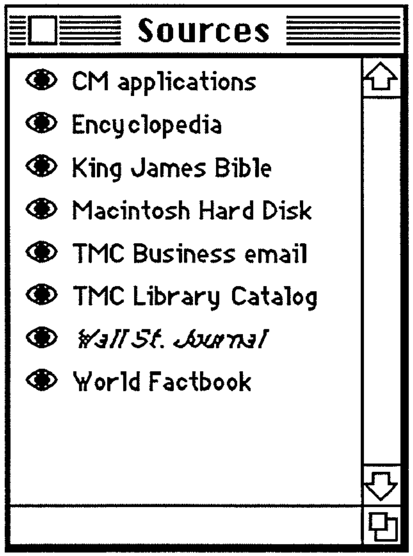
} see Stanfill and Kahle
Communications of the ACM
December 1986

- << 1 sec. response

- Much larger DBs searchable with CM-5
  and inverted index algorithms: 100s to 1000s of Gigabytes

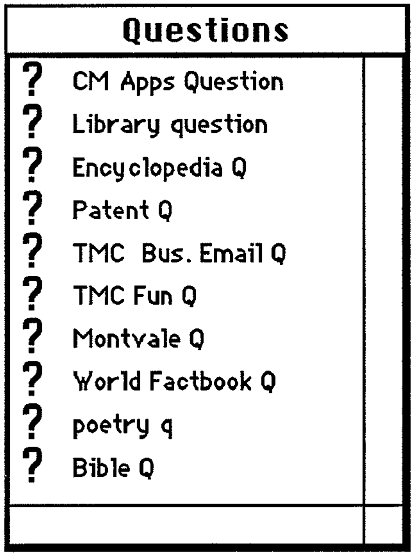
WAIStation

WAIStation: active database sources, saved Questions



Sources

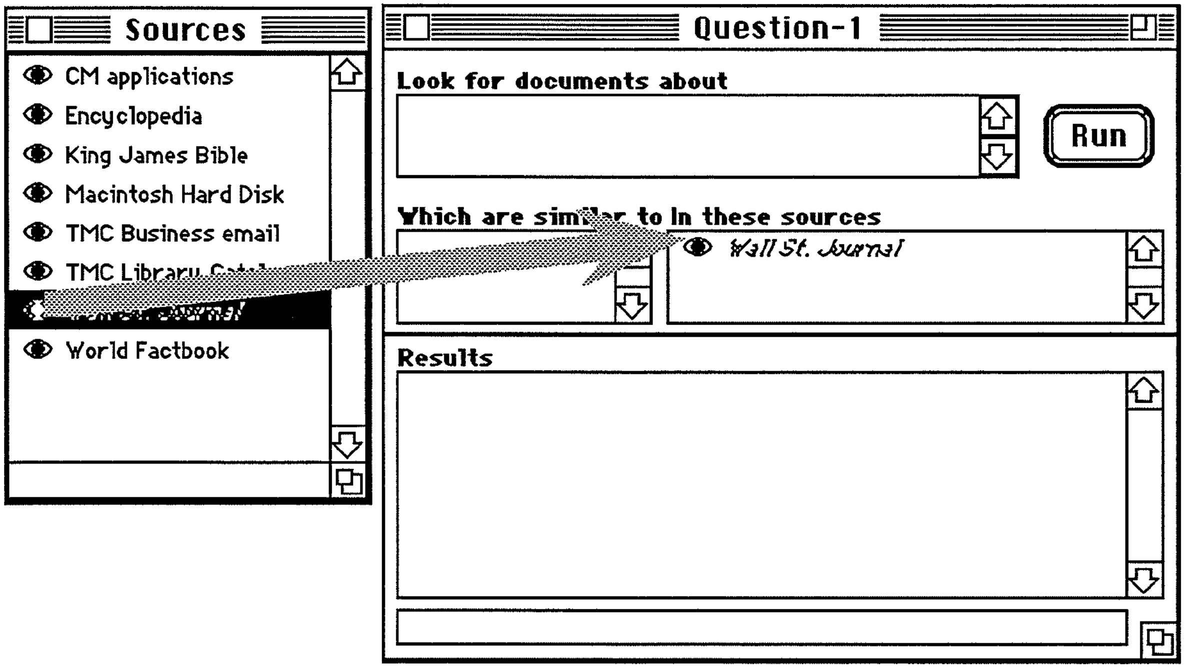
- 👁 CM applications
- 👁 Encyclopedia
- 👁 King James Bible
- 👁 Macintosh Hard Disk
- 👁 TMC Business email
- 👁 TMC Library Catalog
- 👁 Wall St. Journal
- 👁 World Factbook



Questions

- ? CM Apps Question
- ? Library question
- ? Encyclopedia Q
- ? Patent Q
- ? TMC Bus. Email Q
- ? TMC Fun Q
- ? Montvale Q
- ? World Factbook Q
- ? poetry q
- ? Bible Q

Select Data Source



Sources

- CM applications
- Encyclopedia
- King James Bible
- Macintosh Hard Disk
- TMC Business email
- TMC Library Catal
- World Factbook

Question-1

Look for documents about

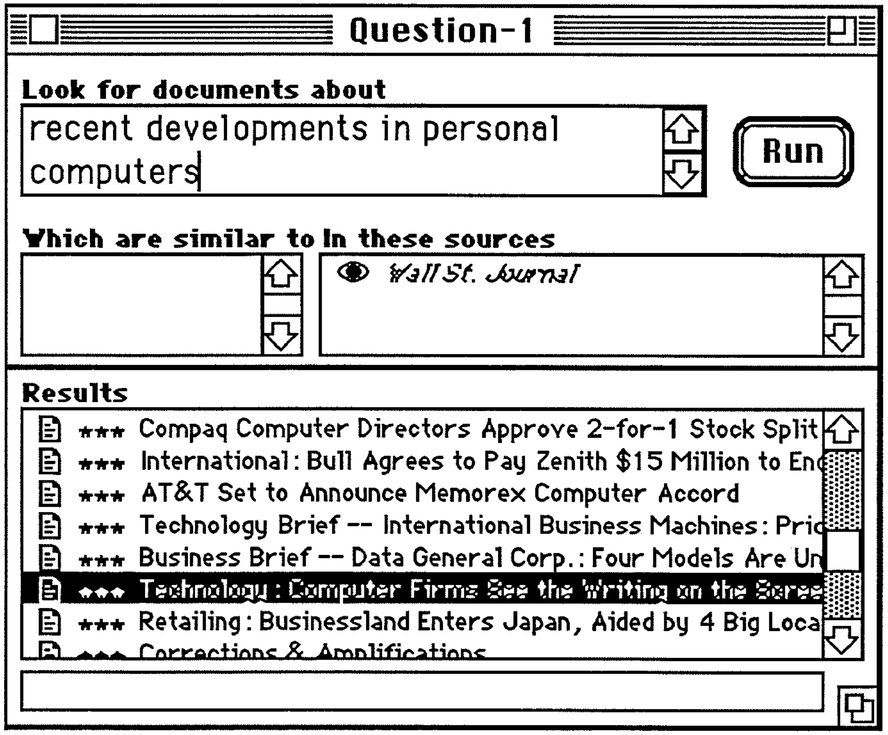
Which are similar to

In these sources

- Wall St. Journal

Results

Run Initial Query



Look for documents about

recent developments in personal computers

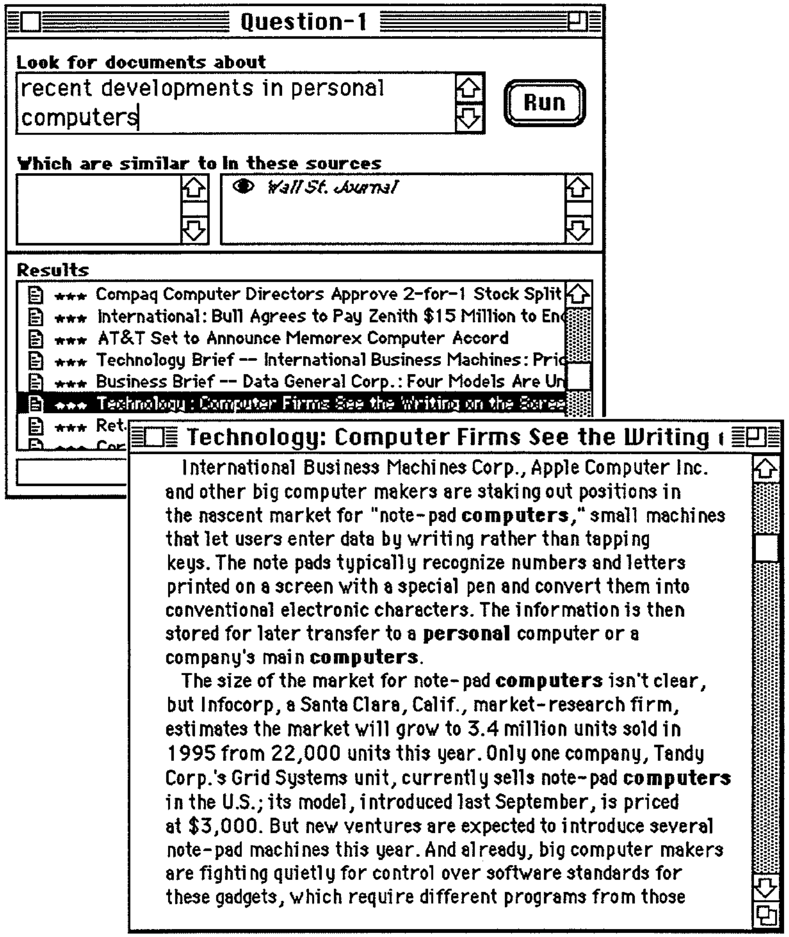
Which are similar to In these sources

Wall St. Journal

Results

- *** Compaq Computer Directors Approve 2-for-1 Stock Split
- *** International: Bull Agrees to Pay Zenith $15 Million to En
- *** AT&T Set to Announce Memorex Computer Accord
- *** Technology Brief -- International Business Machines: Pric
- *** Business Brief -- Data General Corp.: Four Models Are Un
- *** Technology: Computer Firms See the Writing on the Scree
- *** Retailing: Businessland Enters Japan, Aided by 4 Big Loca
- *** Corrections & Amplifications

Click a Headline to Display a Document



Look for documents about

recent developments in personal computers

Which are similar to

In these sources

Wall St. Journal

Results

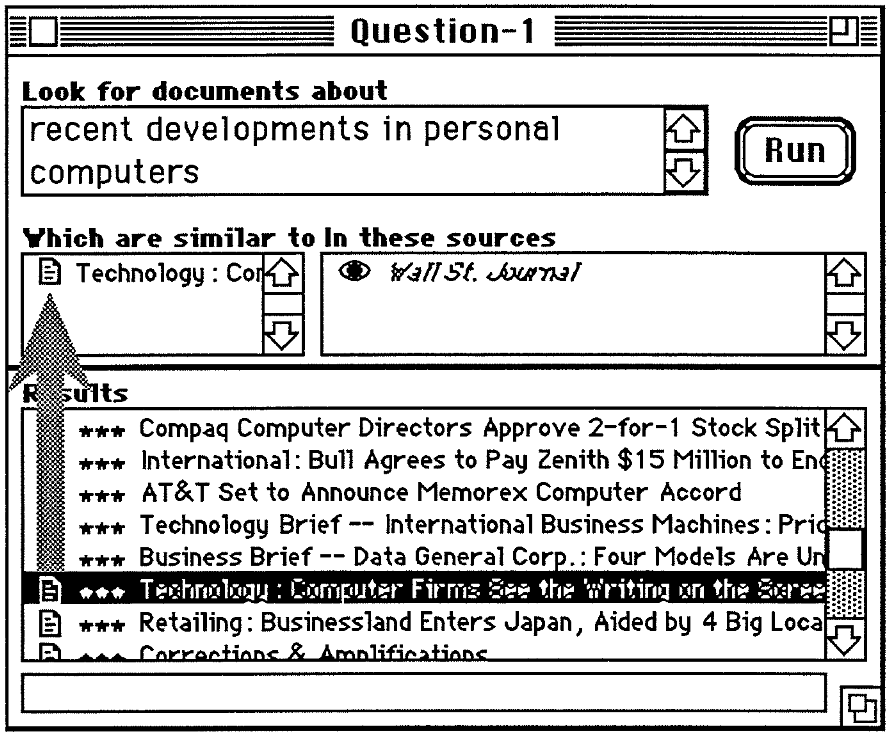
- *** Compaq Computer Directors Approve 2-for-1 Stock Split
- *** International: Bull Agrees to Pay Zenith $15 Million to En
- *** AT&T Set to Announce Memorex Computer Accord
- *** Technology Brief -- International Business Machines: Pri
- *** Business Brief -- Data General Corp.: Four Models Are Un
- *** Technology : Computer Firms See the Writing on the Scree
- *** Ret
- *** Cor

Technology: Computer Firms See the Writing

International Business Machines Corp., Apple Computer Inc. and other big computer makers are staking out positions in the nascent market for "note-pad **computers**," small machines that let users enter data by writing rather than tapping keys. The note pads typically recognize numbers and letters printed on a screen with a special pen and convert them into conventional electronic characters. The information is then stored for later transfer to a **personal** computer or a company's main **computers**.

The size of the market for note-pad **computers** isn't clear, but Infocorp, a Santa Clara, Calif., market-research firm, estimates the market will grow to 3.4 million units sold in 1995 from 22,000 units this year. Only one company, Tandy Corp.'s Grid Systems unit, currently sells note-pad **computers** in the U.S.; its model, introduced last September, is priced at $3,000. But new ventures are expected to introduce several note-pad machines this year. And already, big computer makers are fighting quietly for control over software standards for these gadgets, which require different programs from those

Relevance feedback: “Find me more like this one”



Look for documents about

recent developments in personal computers

Which are similar to

Technology : Co

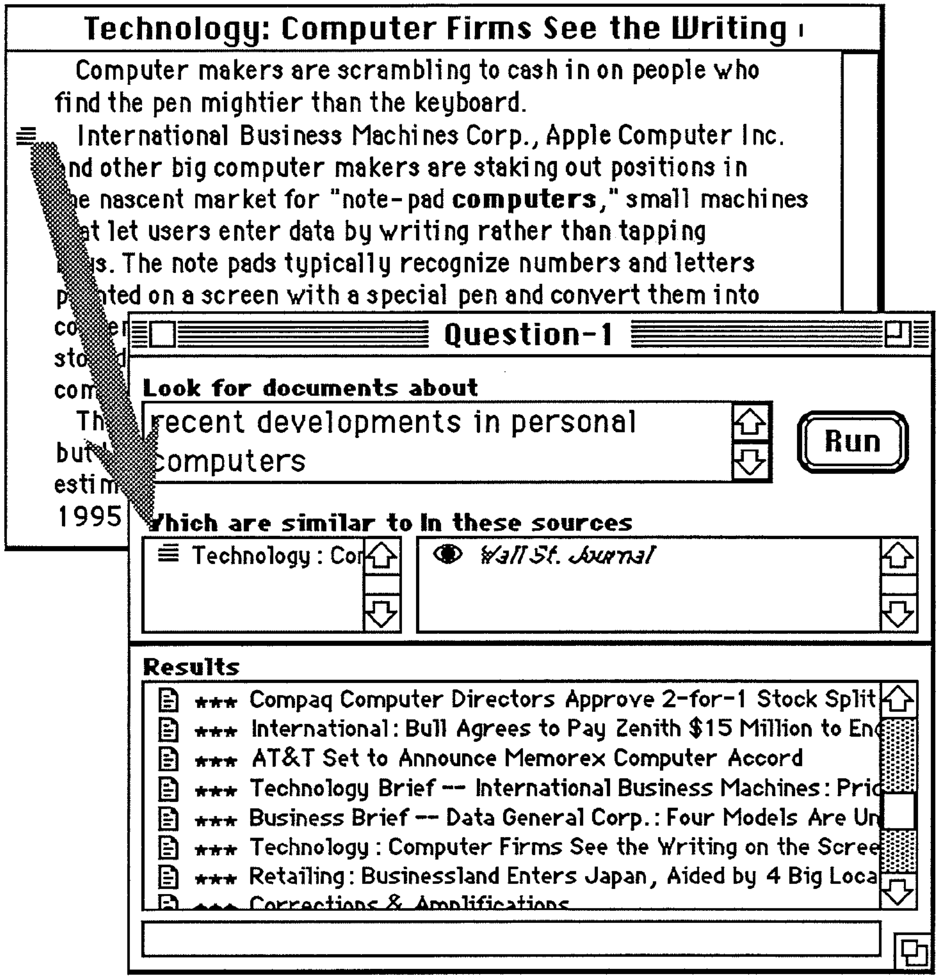
In these sources

Wall St. Journal

Results

- *** Compaq Computer Directors Approve 2-for-1 Stock Split
- *** International: Bull Agrees to Pay Zenith $15 Million to En
- *** AT&T Set to Announce Memorex Computer Accord
- *** Technology Brief -- International Business Machines: Pri
- *** Business Brief -- Data General Corp.: Four Models Are Un
- *** Technology : Computer Firms See the Writing on the Scree
- *** Retailing: Businessland Enters Japan, Aided by 4 Big Loca
- *** Corrections & Amplifications

Relevance Feedback of Paragraph



Technology: Computer Firms See the Writing

Computer makers are scrambling to cash in on people who find the pen mightier than the keyboard.

International Business Machines Corp., Apple Computer Inc.
nd other big computer makers are staking out positions in
e nascent market for "note-pad **computers**," small machines
at let users enter data by writing rather than tapping
s. The note pads typically recognize numbers and letters
p ted on a screen with a special pen and convert them into
co e
sto
com
Th
but
esti
1995

Question-1

Look for documents about

recent developments in personal computers

Run

Which are similar to

≡ Technology : Co

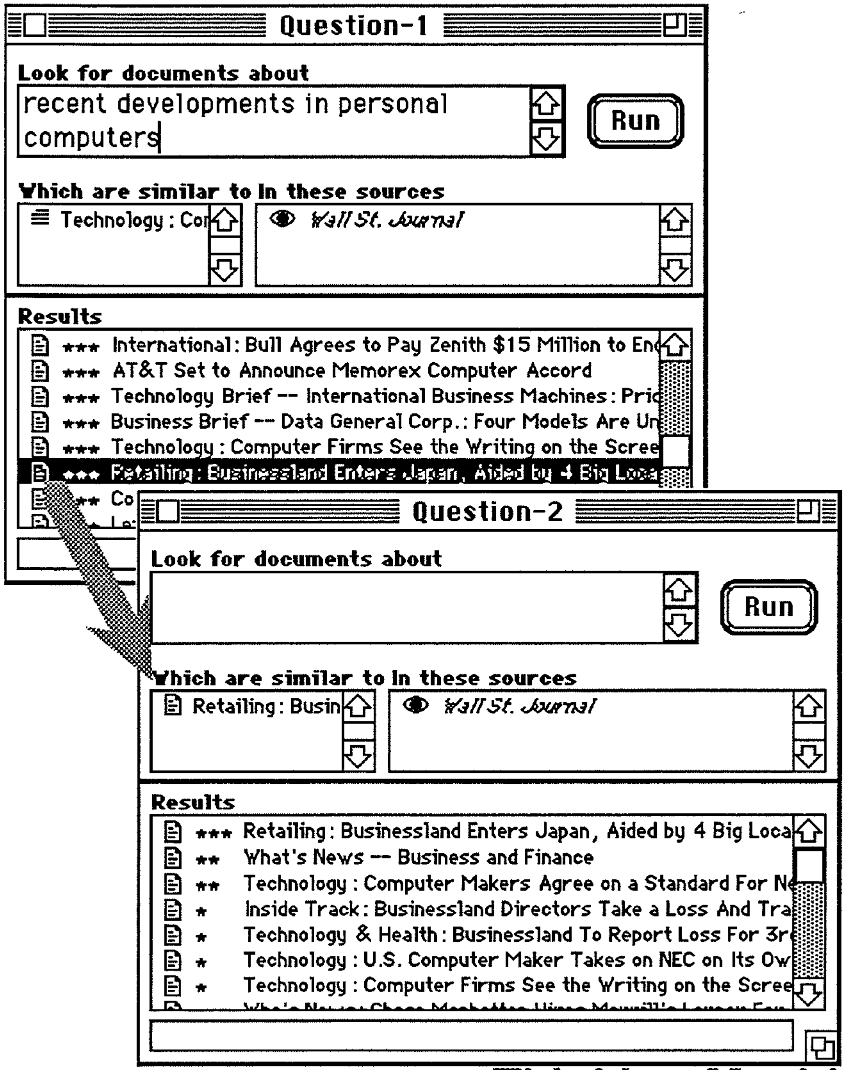
In these sources

Wall St. Journal

Results

- ★★★ Compaq Computer Directors Approve 2-for-1 Stock Split
- ★★★ International: Bull Agrees to Pay Zenith $15 Million to En
- ★★★ AT&T Set to Announce Memorex Computer Accord
- ★★★ Technology Brief -- International Business Machines: Pri
- ★★★ Business Brief -- Data General Corp.: Four Models Are Ur
- ★★★ Technology : Computer Firms See the Writing on the Scree
- ★★★ Retailing: Businessland Enters Japan, Aided by 4 Big Loca
- ★★★ Corrections & Amplifications

"Chaining" of Questions to Follow a Tangent



Question-1

Look for documents about

recent developments in personal computers

Which are similar to

Technology : Co

In these sources

Wall St. Journal

Results

- *** International: Bull Agrees to Pay Zenith $15 Million to En
- *** AT&T Set to Announce Memorex Computer Accord
- *** Technology Brief -- International Business Machines: Pri
- *** Business Brief -- Data General Corp.: Four Models Are Un
- *** Technology : Computer Firms See the Writing on the Scree
- *** Retailing : Businessland Enters Japan, Aided by 4 Big Loca
- *** Co

Question-2

Look for documents about

Which are similar to

Retailing: Busin

In these sources

Wall St. Journal

Results

- *** Retailing: Businessland Enters Japan, Aided by 4 Big Loca
- ** What's News -- Business and Finance
- ** Technology : Computer Makers Agree on a Standard For N
- * Inside Track: Businessland Directors Take a Loss And Tra
- * Technology & Health: Businessland To Report Loss For 3r
- * Technology : U.S. Computer Maker Takes on NEC on Its Ow
- * Technology : Computer Firms See the Writing on the Scree

TMC Internet Release

- CM product for TCP/IP (complete server)
- Example User interfaces for free (no support)
  Macintosh, Gnu Emacs, Xwindows
- Example unix server software to create servers
- Directory of Servers on the internet at least through '91
- 160 Servers now: Weather Maps, patents, journal abstracts, email archives, usenet recipies,...
- Free Software via FTP from Think.com:`/wais/*`
  Mailing list: `wais-discussion-request@think.com`

WAIS Uses

- Over 10,000 users on the Internet
- Users in 24 Countries: Mexico, Singapore, Finland, Australia, etc
- 160 Databases served from 9 Countries: Norway, Canada, UK, etc.
  Average 3 new databases registered per week.

WAIS Uses:
Campus Wide Info Servers

- Class catalog and schedule
- Campus events: movies, sports
- Job listings
- Library catalog
- Phone book
- Professor research interests
- Past theses

```
[     sol.acs.unt.edu]  UNTComputerDoc
[       xantos.uio.no]  UiO_Publications
[   next2.oit.unc.edu]  ibm.pc.FAQ
```

WAIS Uses: Libraries

- Easy to use card catalog
- Remote use from home or office
- Pictures, full text, scanned documents

```
[pegun.law.columbia.e]  columbia-law-library-catalog
[pegun.law.columbia.e]  columbia-spanish-law-catalog
[     quake.think.com]  tmc-library
```

WAIS Uses: Biology

- Journal Abstracts
- Sequence archives
- Images

Currently over 20 Biology databases in Finland, Netherlands, and US

```
[      cmns.think.com]  Molecular-biology
[          bio.vu.nl]  biology-compounds
[     genbank.bio.net]  biology-journal-contents
[       wais.funet.fi]  bionic-ai-researchers
[       wais.funet.fi]  bionic-directory-of-servers
[       wais.funet.fi]  bionic-enzyme
```

WAIS Uses: Chemistry CORE Project

- All published chemistry (8 years all ACS)
- Scanned pictures, ascii text
- Optical jukebox mass storage
- Connection Machine / Newton search engines

Project of :Bellcore, ACS, Chem Abstracts, OCLC, Cornell, and Thinking Machines

```
[  cujo.curtin.edu.au]  chem-eng-current-contents
```

WAIS Uses: Business Executives

- Dow Jones information
- Corporate information
- Personal information

Project: KPMG, Apple, Thinking Machines, Dow Jones

```
[      cmns.think.com]  wall-street-journal-sample
[           think.com]  Business-email
```

WAIS Uses: Medical Researchers/Doctors

- Medical papers
- Storing and matching patient records
- Remote connections to specialized databases

```
[          wais.funet.fi]  bionic-databases-limb
```

WAIS Uses: Community Information

- Dial-up users: no network required
- Directories of services or facilities
- Education and entertainment

```
[      quake.think.com]  internet-resource-guide
[      sol.acs.unt.edu]  online-libraries
[      quake.think.com]  weather
[ lambada.oit.unc.edu]  nsf-bulletins
```

Conclusion

- Electronic Publishing can fill niches now
- Companies are positioning themselves now (workstations, server, and info providers)
- Thinking Machines is the "Engine of the Information Industry"

Wide Area Information Servers: A Supercomputer on every Desk

Brewster Kahle
Thinking Machines Corporation

Wide Area Information Servers: A Supercomputer on every Desk

Brewster Kahle
Thinking Machines Corporation

What I really want...

- My personal information to be accessible
- Published information should find me
- Usable anywhere
- Others can use what I have learned (if I want them to)

What I really want...

- My personal information to be accessible
- Published information should find me
- Usable anywhere
- Others can use what I have learned (if I want them to)

What is it?

Electronic Publishing

(Or publishing over wires)

What is it?

Electronic Publishing

(Or publishing over wires)

New Communications Technology Problems

	BOOKS	Telegraph> Telephone	Electronic Publishing
Experts only	*Monks*	*Operators*	*Professional searchers*
Distribution is hard and expensive	*Vellum is calf skin*	*Telephones on barb wire*	*$1/minute over obscure modems*
Different interfaces	*1000's of languages in Europe alone*	*Switching was manual*	*//query (W5) inform?*
Material is intractable	*Scrolls and manuscripts were about as random access as musical scores*	*No white pages*	*600 databases on Dialog* *~1 Terabyte* *140Gbyte at DJ* *80GB card catalog at RLG*
Business model needed	*Centralized printing*	*Pay per minute*	*Not understood*

New Communications Technology Problems

	BOOKS	Telegraph> Telephone	Electronic Publishing
Experts only	*Monks*	*Operators*	*Professional searchers*
Distribution is hard and expensive	*Vellum is calf skin*	*Telephones on barb wire*	*$1/minute over obscure modems*
Different interfaces	*1000's of languages in Europe alone*	*Switching was manual*	*//query (W5) inform?*
Material is intractable	*Scrolls and manuscripts were about as random access as musical scores*	*No white pages*	*600 databases on Dialog* *~1 Terabyte* *140Gbyte at DJ* *80GB card catalog at RLG*
Business model needed	*Centralized printing*	*Pay per minute*	*Not understood*

Navigation Techniques: Paper

- Alphabetical Listings (dictionary, Encyclopedia)
- Indices (back of the book and Readers Guide)
- Table of Contents (outlining)
- Citation index
- "Tree of Knowledge"
- Have you read any good books lately?

Navigation Techniques: Paper

- Alphabetical Listings (dictionary, Encyclopedia)
- Indices (back of the book and Readers Guide)
- Table of Contents (outlining)
- Citation index
- "Tree of Knowledge"
- Have you read any good books lately?

Navigation Techniques: Computers

- Hierarchical File Systems
- Unix "find" and "grep", Mac "find file"
- Boolean query systems (...within 5 words of...)
- Static Hypertext links (see also pointers)

Navigation Techniques: Computers

- Hierarchical File Systems
- Unix "find" and "grep", Mac "find file"
- Boolean query systems (...within 5 words of...)
- Static Hypertext links (see also pointers)

Navigation Techniques: WAIS

- English language questions and Relevance feedback
  * Iterative retrieval
  * Question-answer dialog
  * Similar to the Newspapers front page the: "continued on page 5"
  * Dynamic Hypertext Links

- 2 level search:
  * Directory of servers (server like any other)
  * Servers themselves

- Copy editors help select documents
  * Easy to "publish" opinions on documents

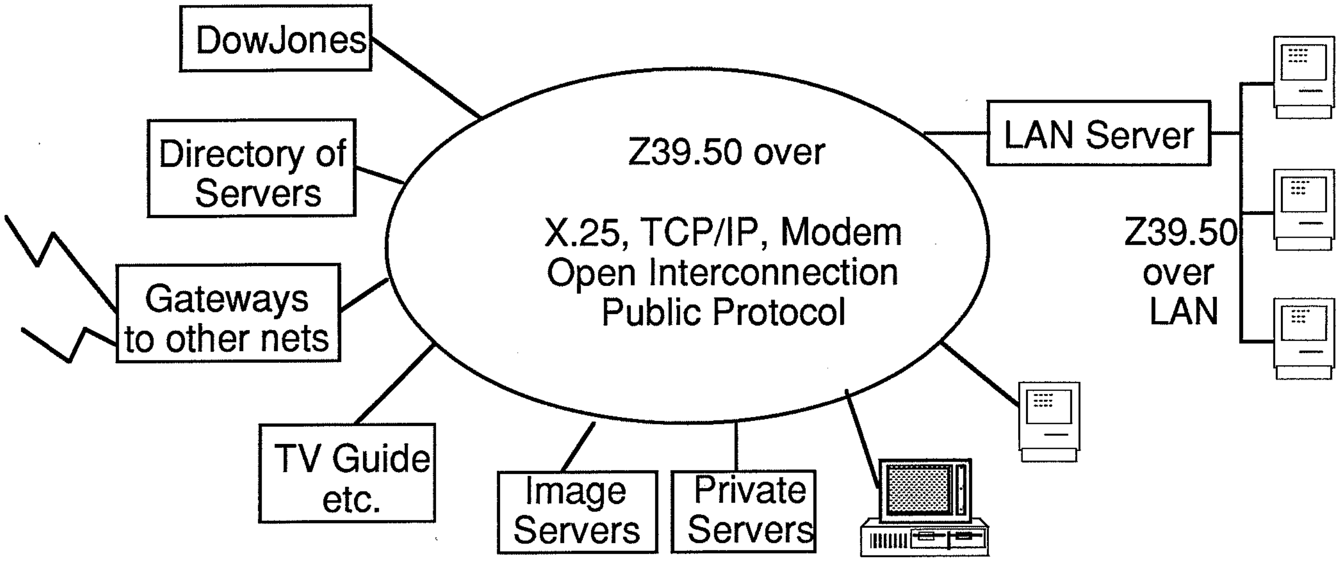
Navigation Techniques: WAIS

- English language questions and Relevance feedback
  * Iterative retrieval
  * Question-answer dialog
  * Similar to the Newspapers front page the: "continued on page 5"
  * Dynamic Hypertext Links

- 2 level search:
  * Directory of servers (server like any other)
  * Servers themselves

- Copy editors help select documents
  * Easy to "publish" opinions on documents

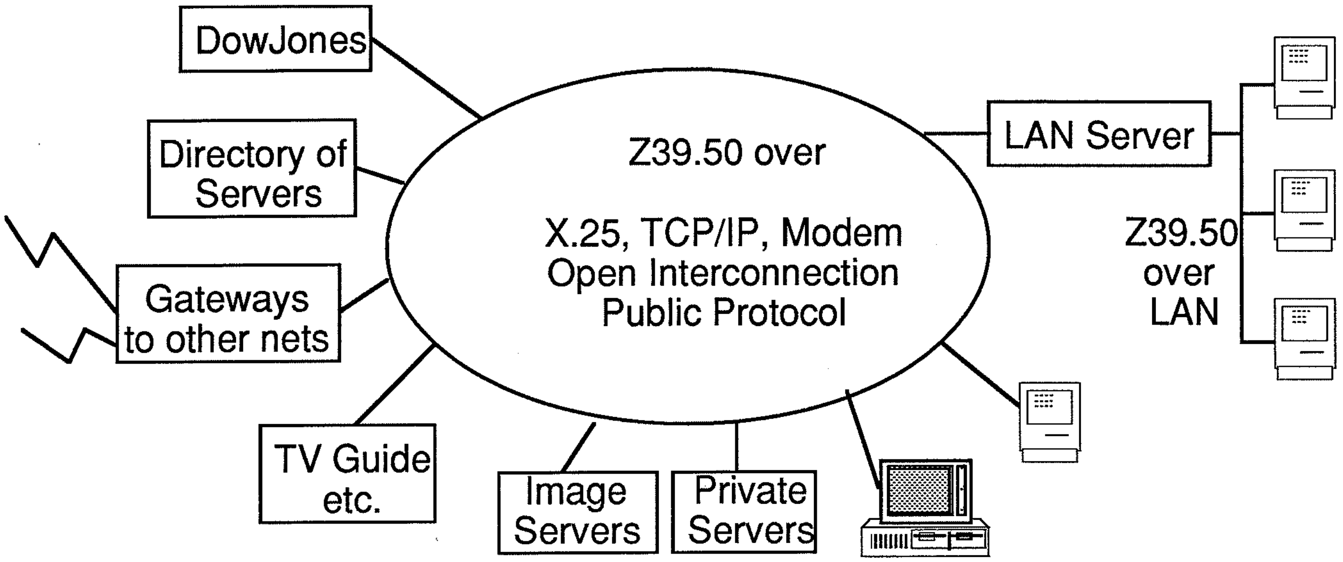
Wide Area Information Server Architecture



Users Needs:
- **Selecting Servers**
- **Answering Questions**
- **Organizing Responses**

Architecture Issues:
- **Scalability**
- **Security**
- **Business model for servers**
- **Reliable Access**

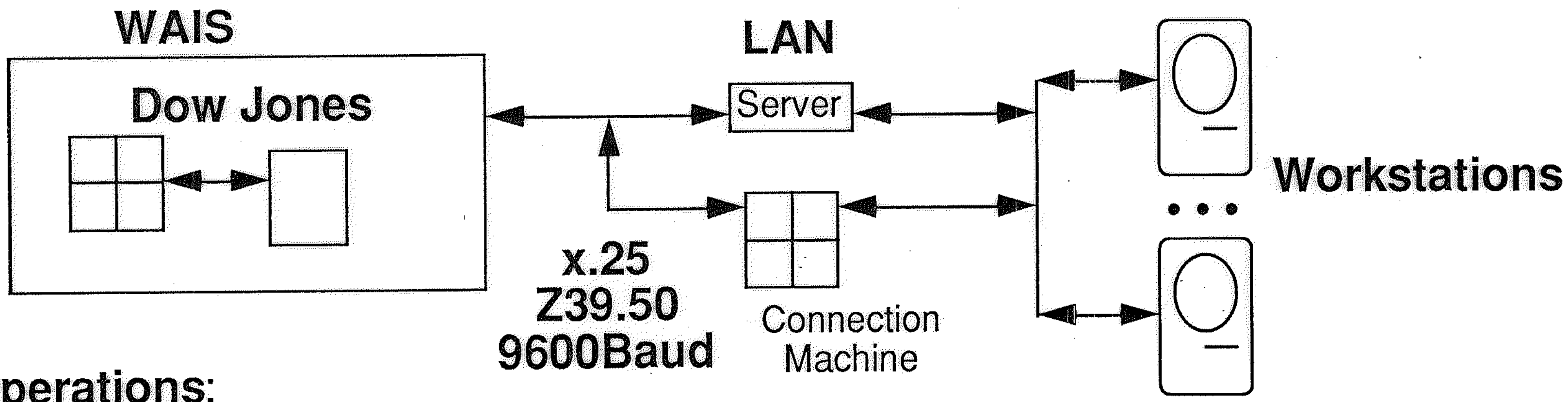
Wide Area Information Server Architecture



Users Needs:
- **Selecting Servers**
- **Answering Questions**
- **Organizing Responses**

Architecture Issues:
- **Scalability**
- **Security**
- **Business model for servers**
- **Reliable Access**

Demonstration System Structure



Operations:
- Archiving
- Queries
- Retrieval

IR Type:
- Broadcast
- Query by Example

Databases:
- Wall St Journal
- Barron's
- 400 Business Mags

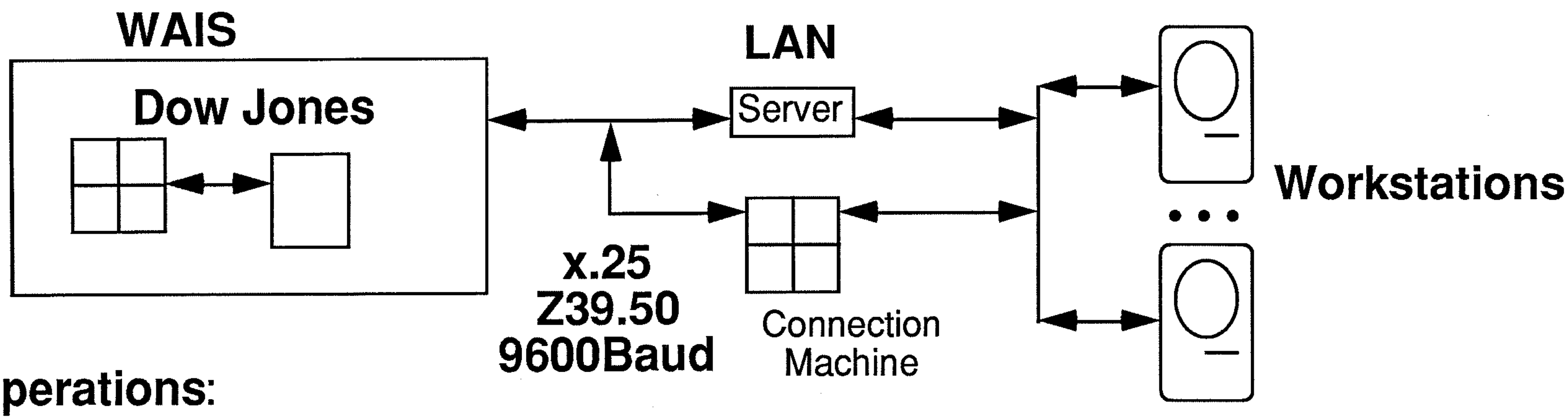
CM: Operations: Queries
IR Type:
enhanced relevance feedback
DBs: DowVision and
memo's, mail,
word processor files

Mac:
Operations:
- Human Int
- Retrieval
- Queries
- "Caching" Docs
- User Profiles

IR Type:
- Query by example

DBs:
- Personal Text
- Cached data

Demonstration System Structure



WAIS

Dow Jones

LAN

Server

Workstations

x.25
Z39.50
9600Baud

Connection
Machine

Operations:
- Archiving
- Queries
- Retrieval

IR Type:
- Broadcast
- Query by Example

Databases:
- Wall St Journal
- Barron's
- 400 Business Mags

CM: **Operations**: Queries
IR Type:
enhanced relevance feedback
DBs: DowVision and
memo's, mail,
word processor files

Mac:
Operations:
- Human Int
- Retrieval
- Queries
- "Caching" Docs
- User Profiles

IR Type:
- Query by example

DBs:
- Personal Text
- Cached data

WAIS Clients

- Busy 24 hours a day finding information
- Ponder all indications of the preferences of its user
- Gossip with other clients about their discoveries
- Scours the world (within a budget) to find new sources

WAIS Clients

- Busy 24 hours a day finding information
- Ponder all indications of the preferences of its user
- Gossip with other clients about their discoveries
- Scours the world (within a budget) to find new sources

WAIS Protocol

- Based on Z39.50, bypass proprietary period
- Flexible
- Non Threatening for corporations

- Search: (words, doc_ids, databases) -> server
  returns list of: (headline, score, doc_id, types)'s
- Retrieval: (doc_id, type, start, end) -> server
  returns: bunch of bytes

- Doc_id: An ISBN for the Electronic Age
  ((orig_server, orig_database, orig_local_id)
  (dist_server, dist_database, dist_local_id)
- Server Description:
  (:ip-address, :database-name, :cost, :description)

WAIS Protocol

- Based on Z39.50, bypass proprietary period
- Flexible
- Non Threatening for corporations

- Search: (words, doc_ids, databases) -> server
  returns list of: (headline, score, doc_id, types)'s
- Retrieval: (doc_id, type, start, end) -> server
  returns: bunch of bytes

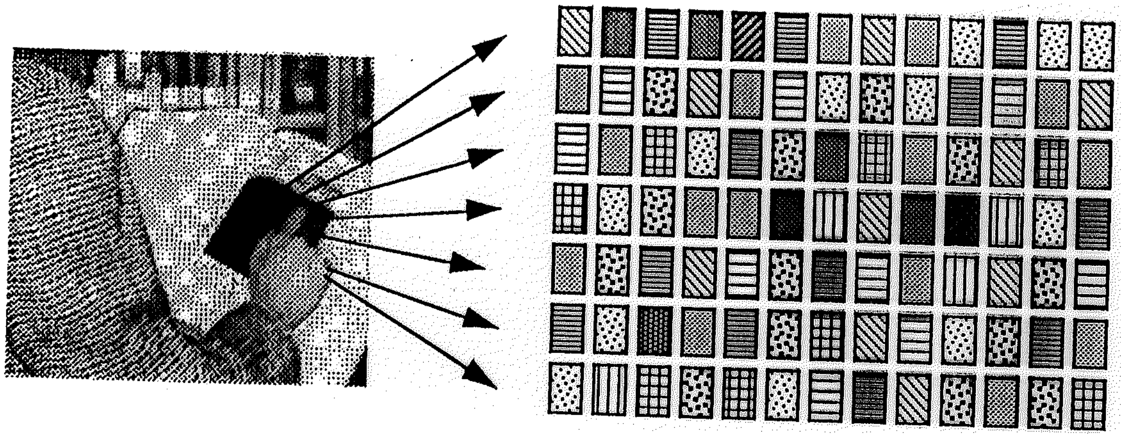
- Doc_id: An ISBN for the Electronic Age
  ((orig_server, orig_database, orig_local_id)
  (dist_server, dist_database, dist_local_id)
- Server Description:
  (:ip-address, :database-name, :cost, :description)

Connection Machine Server

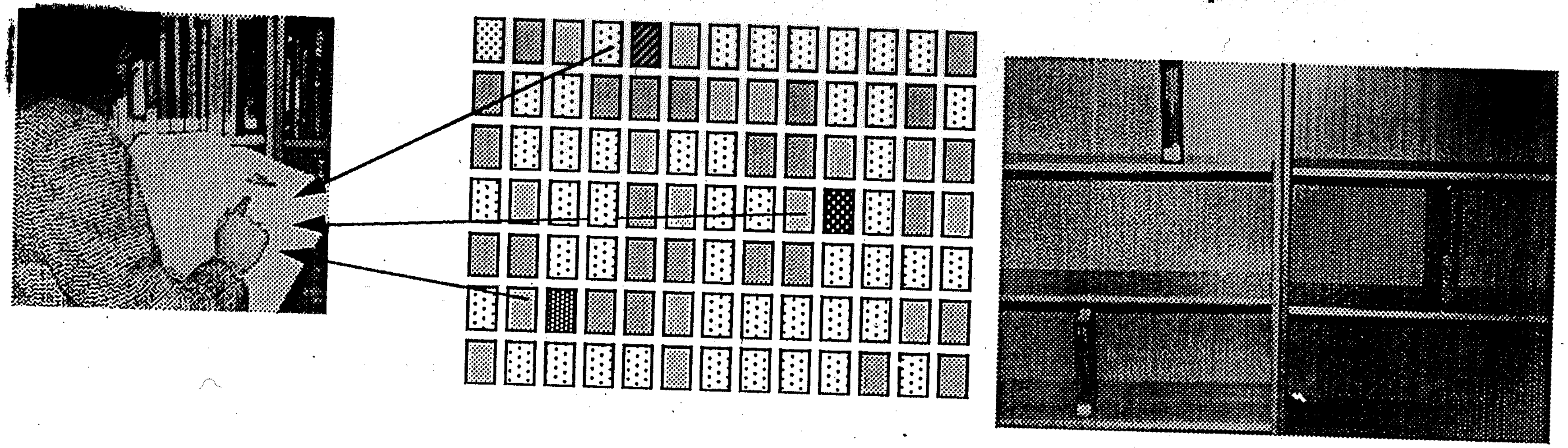
- 1-25GBytes (and getting bigger)
- Supports thousands of users
- Automatic Indexing
- Uses words and phrases in question to find appropriate documents
- First turn-key massively parallel application

Connection Machine Server

- 1-25GBytes (and getting bigger)
- Supports thousands of users
- Automatic Indexing
- Uses words and phrases in question to find appropriate documents
- First turn-key massively parallel application



An example document is compared to all the others, in parallel.



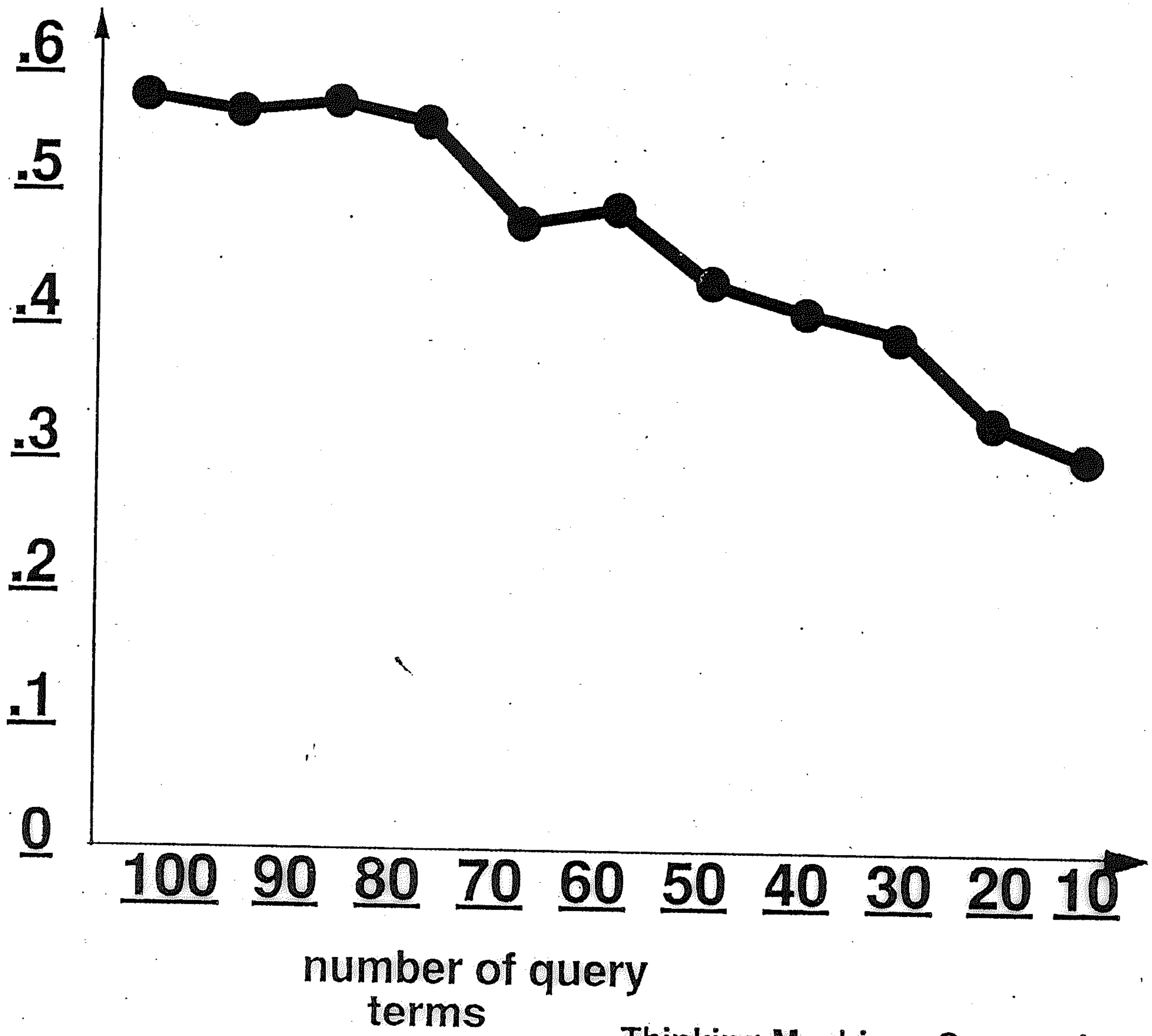
Only the best matches are presented to the user.

Connection Machine Server

- 1-25GBytes (and getting bigger)
- Supports thousands of users
- Automatic Indexing
- Uses words and phrases in question to find appropriate documents
- First turn-key massively parallel application

Results Improve with Query Size

Precision x
recall
@ 25% recall



Average performance over 13 reference sets

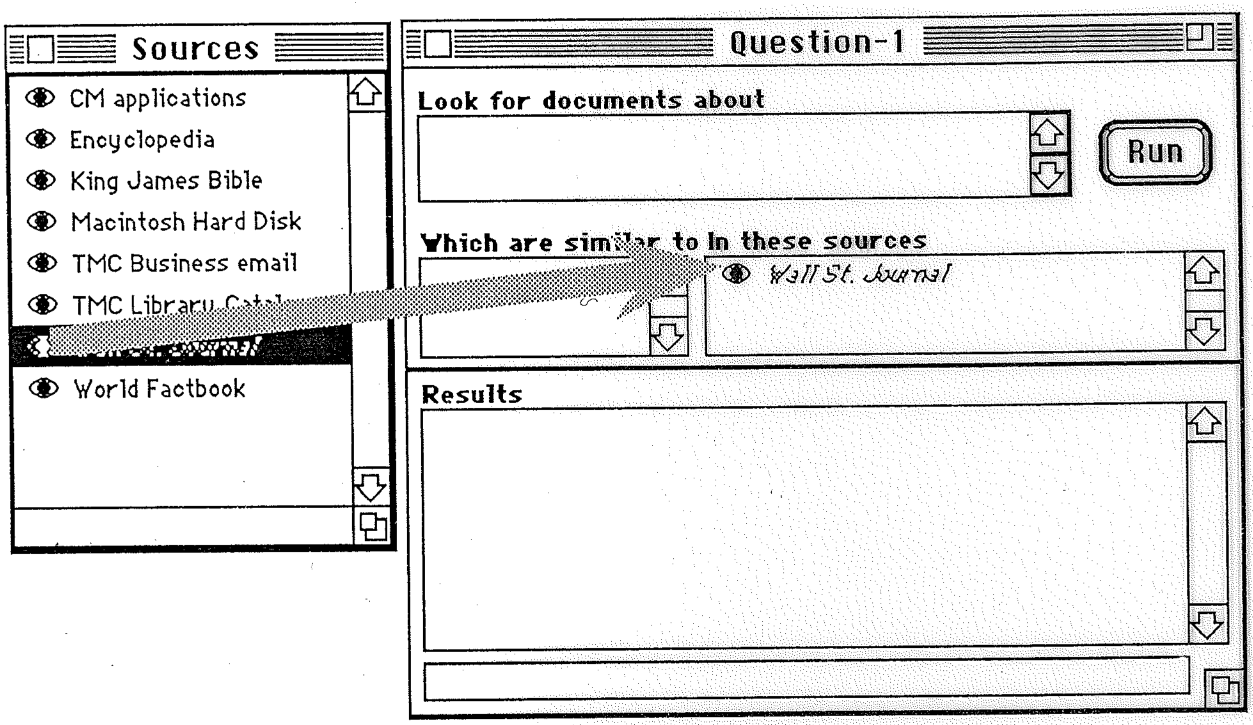
How Fast?

10–term query

DB Size	Procs	DVs	Time	Storage Method
1.5 GB	4K	0	0.055	Main Memory
3 GB	8K	0	0.055	Main Memory
6 GB	16K	0	0.055	Main Memory
12 GB	32K	0	0.055	Main Memory
24 GB	64K	0	0.055	Main Memory
64 GB	8K	1	1.7	Independent Disk
128 GB	8K	1	2.8	Independent Disk
256 GB	16K	2	3.6	Striped Disk
512 GB	32K	4	3.6	Striped Disk
1024 GB	64K	8	3.6	Striped Disk
2048 GB	64K	16	5.1	Striped Disk
4096 GB	64K	32	8.2	Striped Disk
8192 GB	64K	64	12.4	Striped Disk

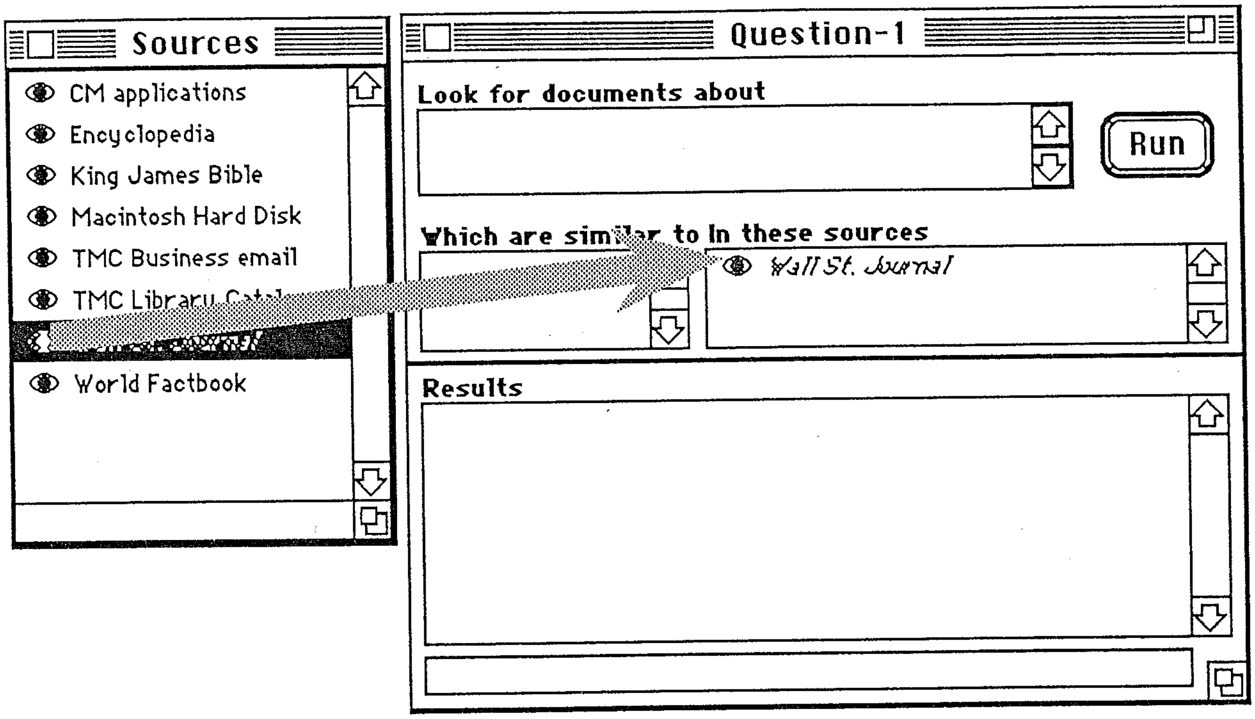
Estimates based on synthetic database, benchmark code.

WAIStation
Step 1



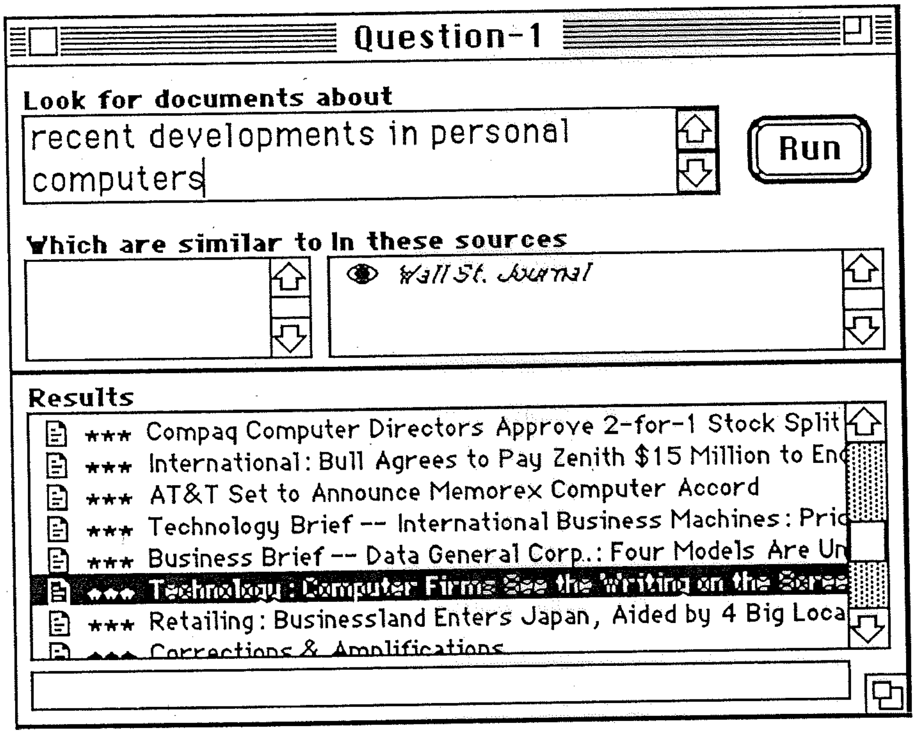
Step 1: Sources are dragged with the mouse into the Question Window. A question can contain multiple sources. When the question is run, it asks for information from each included source.

WAIStation
Step 1



Step 1: Sources are dragged with the mouse into the Question Window. A question can contain multiple sources. When the question is run, it asks for information from each included source.

WAIStation
Step 2



Question-1

Look for documents about
recent developments in personal computers

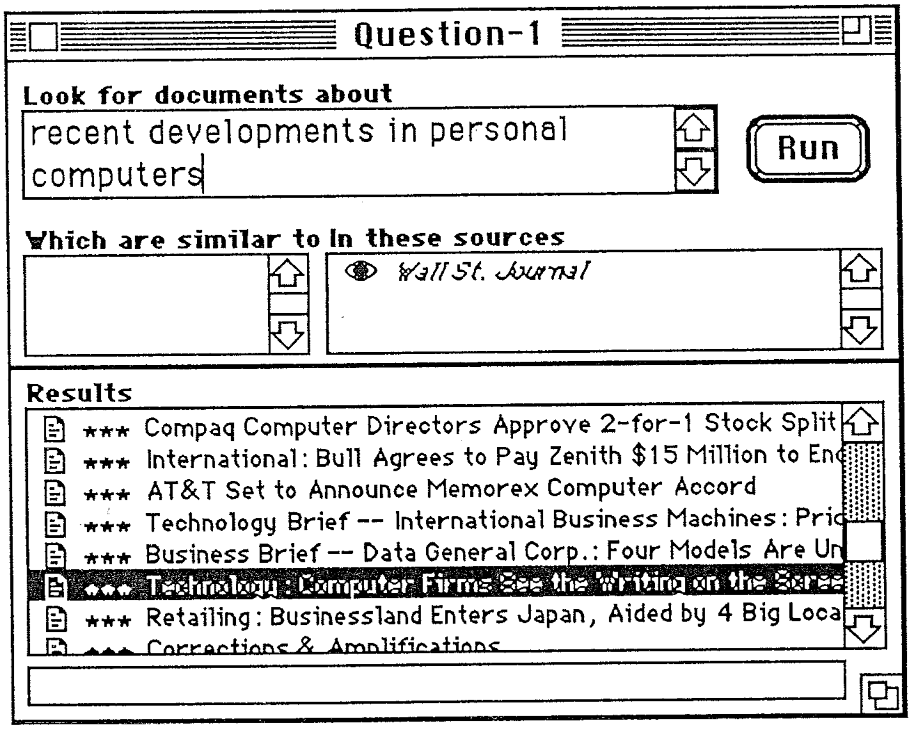
Vhich are similar to

In these sources
Wall St. Journal

Results
- *** Compaq Computer Directors Approve 2-for-1 Stock Split
- *** International: Bull Agrees to Pay Zenith $15 Million to En
- *** AT&T Set to Announce Memorex Computer Accord
- *** Technology Brief -- International Business Machines: Pri
- *** Business Brief -- Data General Corp..: Four Models Are U
- *** Technology: Computer Firms See the Writing on the Scree
- *** Retailing: Businessland Enters Japan, Aided by 4 Big Loca
- *** Corrections & Amplifications

Step 2: When a query is run, headlines of documents satisfying the query are displayed.

WAIStation
Step 2



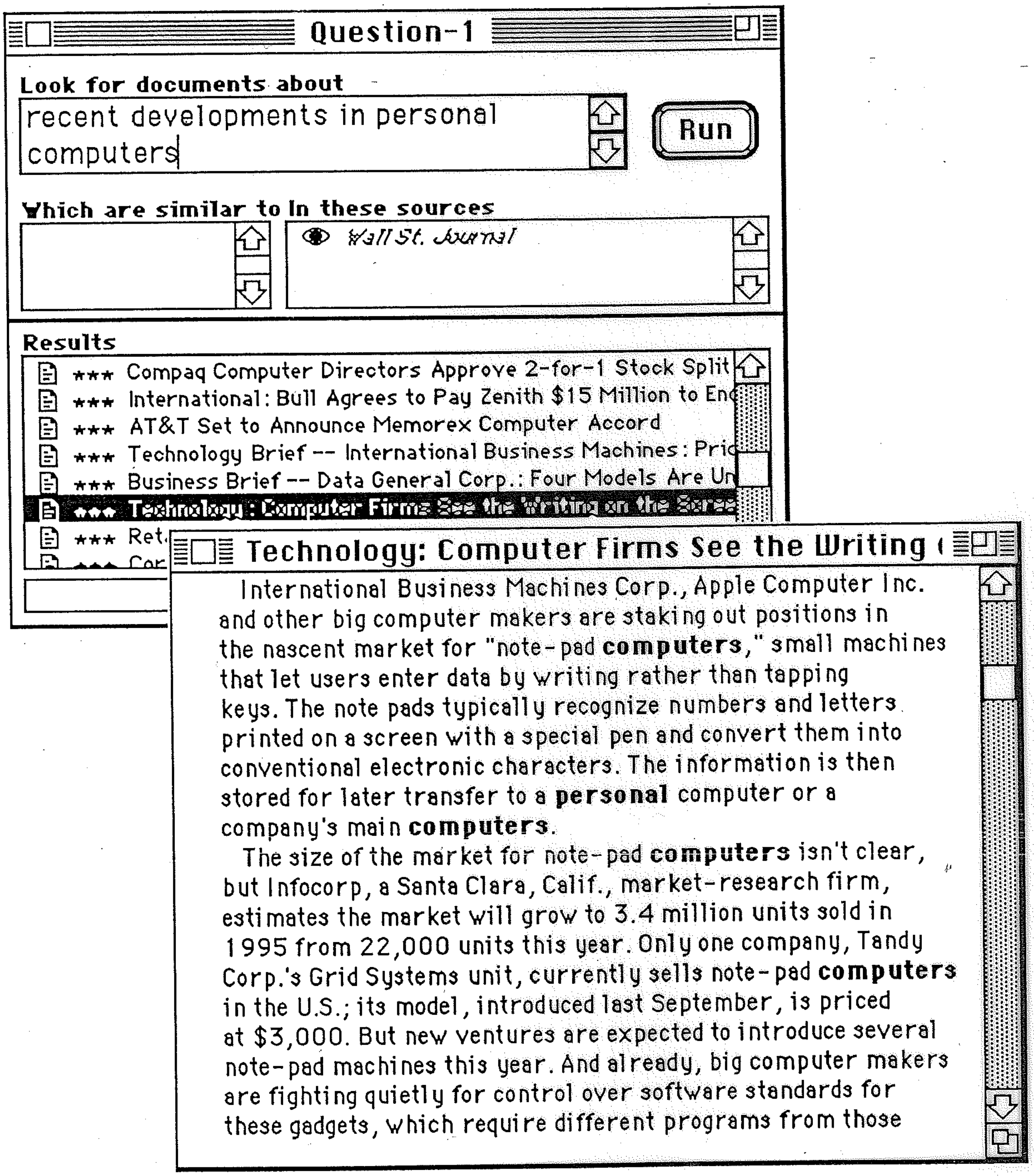
Look for documents about
recent developments in personal computers

Which are similar to
In these sources
Wall St. Journal

Results
*** Compaq Computer Directors Approve 2-for-1 Stock Split
*** International: Bull Agrees to Pay Zenith $15 Million to End
*** AT&T Set to Announce Memorex Computer Accord
*** Technology Brief -- International Business Machines: Pric
*** Business Brief -- Data General Corp.: Four Models Are Un
*** Technology: Computer Firms See the Writing on the Scree
*** Retailing: Businessland Enters Japan, Aided by 4 Big Loca
*** Corrections & Amplifications

Step 2: When a query is run, headlines of documents satisfying the query are displayed.

WAIStation
Step 3



Look for documents about
recent developments in personal computers

Which are similar to

In these sources
Wall St. Journal

Results

*** Compaq Computer Directors Approve 2-for-1 Stock Split
*** International: Bull Agrees to Pay Zenith $15 Million to En
*** AT&T Set to Announce Memorex Computer Accord
*** Technology Brief -- International Business Machines: Pri
*** Business Brief -- Data General Corp.: Four Models Are Un
*** Technology: Computer Firms See the Writing on the Scree
*** Ret
*** Cor

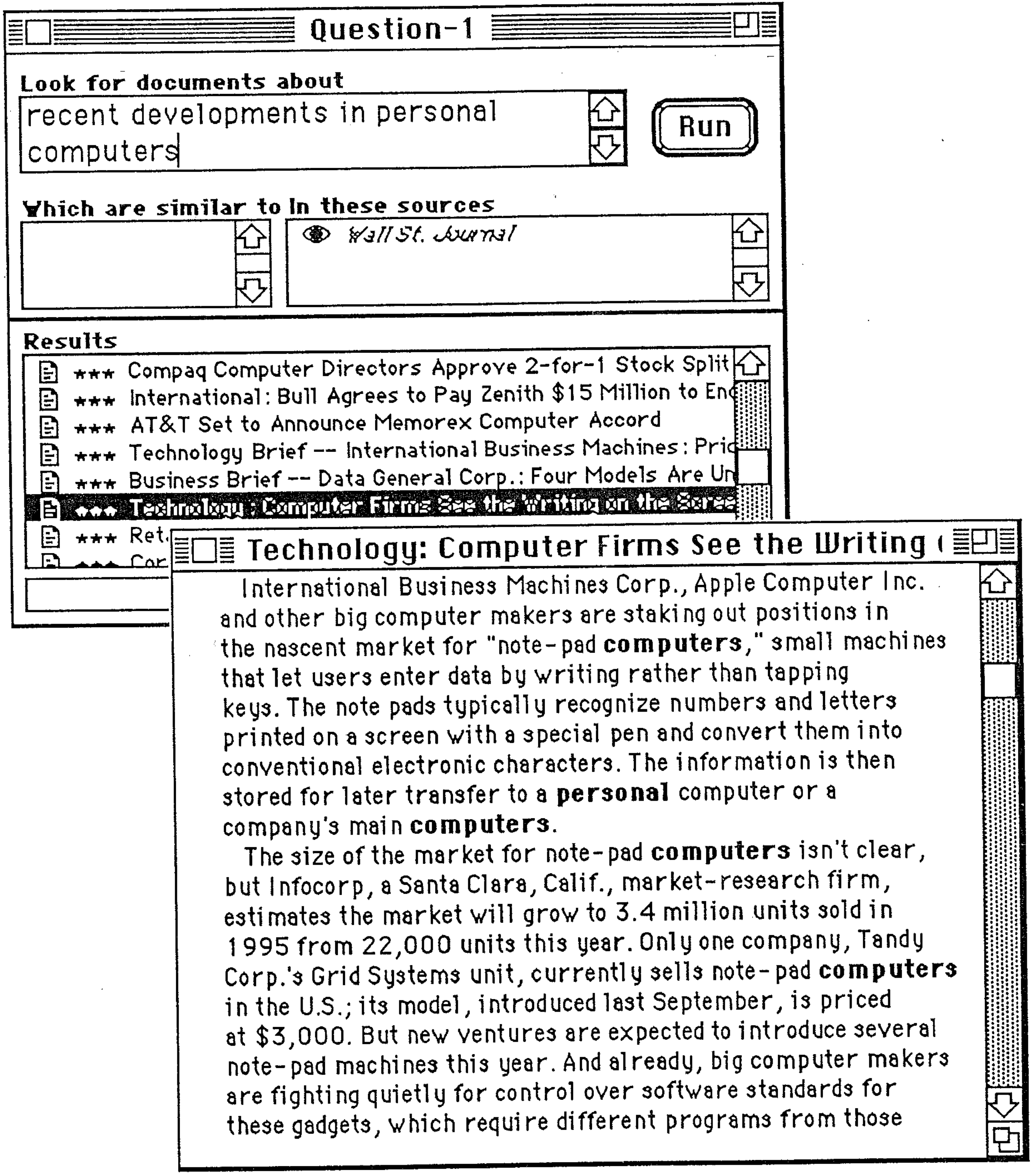
Technology: Computer Firms See the Writing (

International Business Machines Corp., Apple Computer Inc. and other big computer makers are staking out positions in the nascent market for "note-pad **computers**," small machines that let users enter data by writing rather than tapping keys. The note pads typically recognize numbers and letters printed on a screen with a special pen and convert them into conventional electronic characters. The information is then stored for later transfer to a **personal** computer or a company's main **computers**.

The size of the market for note-pad **computers** isn't clear, but Infocorp, a Santa Clara, Calif., market-research firm, estimates the market will grow to 3.4 million units sold in 1995 from 22,000 units this year. Only one company, Tandy Corp.'s Grid Systems unit, currently sells note-pad **computers** in the U.S.; its model, introduced last September, is priced at $3,000. But new ventures are expected to introduce several note-pad machines this year. And already, big computer makers are fighting quietly for control over software standards for these gadgets, which require different programs from those

Step 3: With the mouse, the user clicks on any result document to retrieve it.

WAIStation
Step 3



Look for documents about

recent developments in personal computers

Which are similar to

In these sources

Wall St. Journal

- *** Compaq Computer Directors Approve 2-for-1 Stock Split
- *** International: Bull Agrees to Pay Zenith $15 Million to En
- *** AT&T Set to Announce Memorex Computer Accord
- *** Technology Brief -- International Business Machines: Pri
- *** Business Brief -- Data General Corp.: Four Models Are Un
- *** Technology: Computer Firms See the Writing on the Scre
- *** Ret
- *** Cor

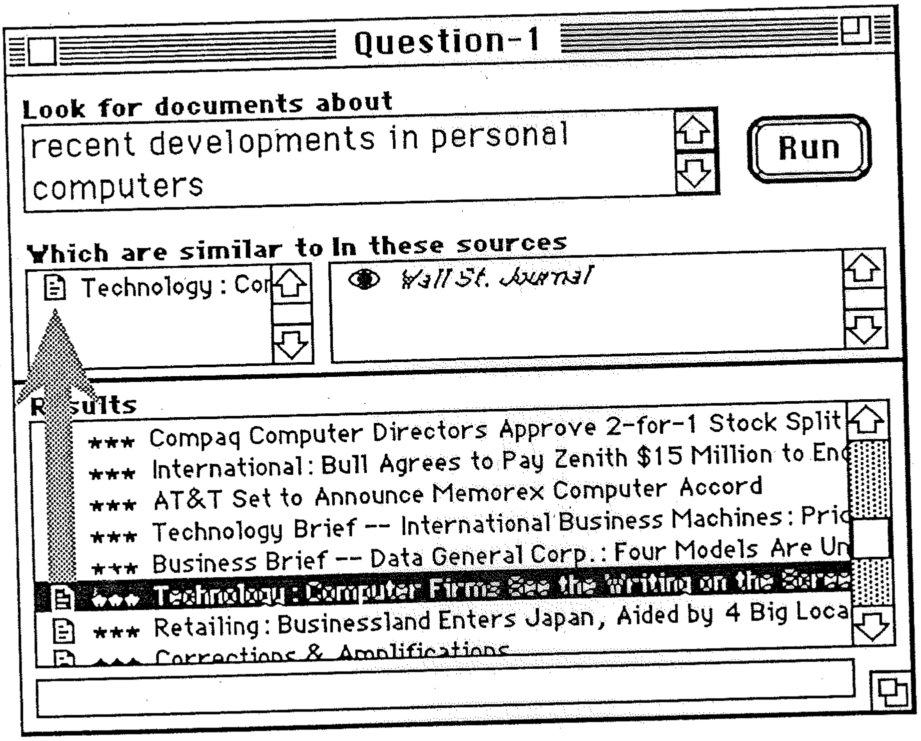
Technology: Computer Firms See the Writing

International Business Machines Corp., Apple Computer Inc. and other big computer makers are staking out positions in the nascent market for "note-pad **computers**," small machines that let users enter data by writing rather than tapping keys. The note pads typically recognize numbers and letters printed on a screen with a special pen and convert them into conventional electronic characters. The information is then stored for later transfer to a **personal** computer or a company's main **computers**.

The size of the market for note-pad **computers** isn't clear, but Infocorp, a Santa Clara, Calif., market-research firm, estimates the market will grow to 3.4 million units sold in 1995 from 22,000 units this year. Only one company, Tandy Corp.'s Grid Systems unit, currently sells note-pad **computers** in the U.S.; its model, introduced last September, is priced at $3,000. But new ventures are expected to introduce several note-pad machines this year. And already, big computer makers are fighting quietly for control over software standards for these gadgets, which require different programs from those

Step 3: With the mouse, the user clicks on any result document to retrieve it.

WAIStation
Step 4



Look for documents about

recent developments in personal computers

Which are similar to

Technology : Co

In these sources

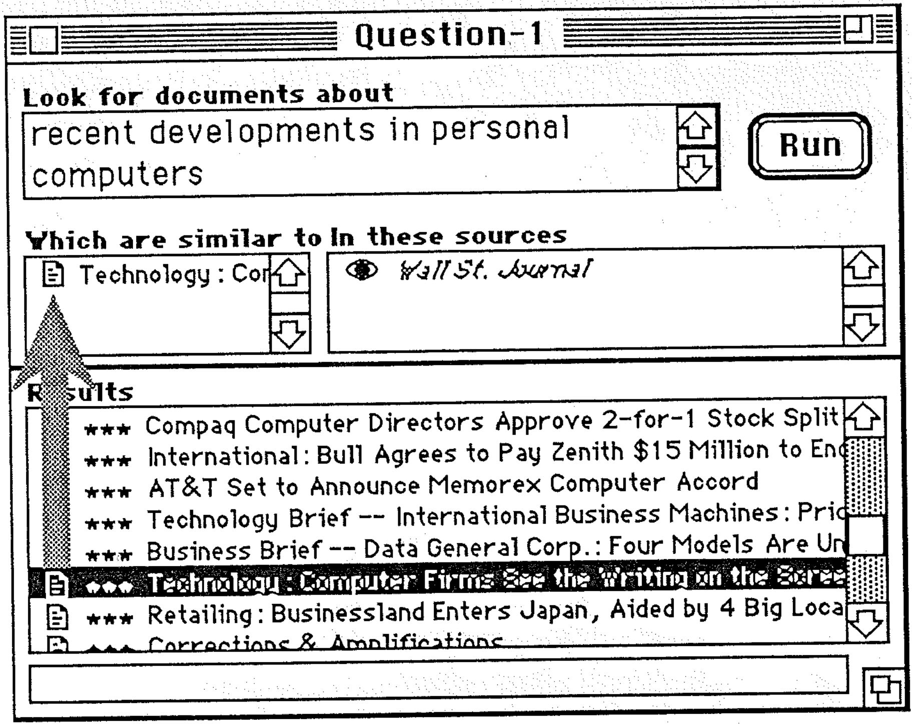
Wall St. Journal

Results

- *** Compaq Computer Directors Approve 2-for-1 Stock Split
- *** International: Bull Agrees to Pay Zenith $15 Million to En
- *** AT&T Set to Announce Memorex Computer Accord
- *** Technology Brief -- International Business Machines: Pri
- *** Business Brief -- Data General Corp.: Four Models Are U
- *** Technology: Computer Firms See the Writing on the Scre
- *** Retailing: Businessland Enters Japan, Aided by 4 Big Loca
- *** Corrections & Amplifications

Step 4: To refine the search, any one or more of the result documents can moved to the "Which are similar to:" box. When the search is run again, the results will be updated to include documents which are "similar" to the ones selected.

WAIStation
Step 4



Look for documents about

recent developments in personal computers

Which are similar to

Technology : Co

In these sources

Wall St. Journal

Results

- *** Compaq Computer Directors Approve 2-for-1 Stock Split
- *** International: Bull Agrees to Pay Zenith $15 Million to En
- *** AT&T Set to Announce Memorex Computer Accord
- *** Technology Brief -- International Business Machines: Pri
- *** Business Brief -- Data General Corp.: Four Models Are Ur
- *** Technology: Computer Firms See the Writing on the Scree
- *** Retailing: Businessland Enters Japan, Aided by 4 Big Loca
- *** Corrections & Amplifications

Step 4: To refine the search, any one or more of the result documents can moved to the "Which are similar to:" box. When the search is run again, the results will be updated to include documents which are "similar" to the ones selected.

TMC Internet Release

- CM product for TCP/IP (complete server)
- Example User interfaces for free (no support)
  Macintosh, Gnu Emacs, Xwindows
- Example unix server software to create servers
- Directory of Servers on the internet at least through '91
- 42 Servers now: Weather Maps, patents, Government programs, Risks-digest, usenet recipies, Lewis Carroll,...
- Anonymous FTP Think.com:/public/wais/*
  Mailing list: wais-discussion-request@think.com

TMC Internet Release

- CM product for TCP/IP (complete server)
- Example User interfaces for free (no support)
  Macintosh, Gnu Emacs, Xwindows
- Example unix server software to create servers
- Directory of Servers on the internet at least through '91
- 42 Servers now: Weather Maps, patents, Government programs, Risks-digest, usenet recipies, Lewis Carroll,...
- Anonymous FTP Think.com:/public/wais/*
  Mailing list: wais-discussion-request@think.com

WAIS Uses: Campus Wide Info Servers

- Class catalog and schedule
- Campus events: movies, sports
- Job listings
- Library catalog
- Phone book
- Professor research interests
- Past theses

```
[     sol.acs.unt.edu]  UNTComputerDoc
[       xantos.uio.no]  UiO_Publications
[   next2.oit.unc.edu]  ibm.pc.FAQ
```

WAIS Uses: Campus Wide Info Servers

- Class catalog and schedule
- Campus events: movies, sports
- Job listings
- Library catalog
- Phone book
- Professor research interests
- Past theses

```
[     sol.acs.unt.edu]  UNTComputerDoc
[        xantos.uio.no]  UiO_Publications
[   next2.oit.unc.edu]  ibm.pc.FAQ
```

WAIS Uses: Libraries

- Easy to use card catalog
- Remote use from home or office
- Pictures, full text, scanned documents

```
[pegun.law.columbia.e]  columbia-law-library-catalog
[pegun.law.columbia.e]  columbia-spanish-law-catalog
[     quake.think.com]  tmc-library
```

WAIS Uses: Libraries

- Easy to use card catalog
- Remote use from home or office
- Pictures, full text, scanned documents

```
[pegun.law.columbia.e]  columbia-law-library-catalog
[pegun.law.columbia.e]  columbia-spanish-law-catalog
[     quake.think.com]  tmc-library
```

WAIS Uses: Biology

- Journal Abstracts
- Sequence archives
- Images

Currently over 20 Biology databases in Finland, Netherlands, and US

```
[       cmns.think.com]  Molecular-biology
[           bio.vu.nl]  biology-compounds
[      genbank.bio.net]  biology-journal-contents
[        wais.funet.fi]  bionic-ai-researchers
[        wais.funet.fi]  bionic-directory-of-servers
[        wais.funet.fi]  bionic-enzyme
```

WAIS Uses: Biology

- Journal Abstracts
- Sequence archives
- Images

Currently over 20 Biology databases in Finland, Netherlands, and US

```
[      cmns.think.com]  Molecular-biology
[           bio.vu.nl]  biology-compounds
[     genbank.bio.net]  biology-journal-contents
[       wais.funet.fi]  bionic-ai-researchers
[       wais.funet.fi]  bionic-directory-of-servers
[       wais.funet.fi]  bionic-enzyme
```

WAIS Uses: Chemistry CORE Project

- All published chemistry (8 years all ACS)
- Scanned pictures, ascii text
- Optical jukebox mass storage
- Connection Machine / Newton search engines

Project of :Bellcore, ACS, Chem Abstracts,
OCLC, Cornell, and Thinking Machines

```
[  cujo.curtin.edu.au]  chem-eng-current-contents
```

WAIS Uses: Chemistry CORE Project

- All published chemistry (8 years all ACS)
- Scanned pictures, ascii text
- Optical jukebox mass storage
- Connection Machine / Newton search engines

Project of :Bellcore, ACS, Chem Abstracts, OCLC, Cornell, and Thinking Machines

```
[  cujo.curtin.edu.au]  chem-eng-current-contents
```

WAIS Uses: Documentation

- Up-to-date documentation
- Online help system
- Distribution of bug notices and fixes
- Mailing list archives

```
CMNS.Think.com  CM-Fortran.src
Quake.think.com wais-talk.src
PRISM CM programming environment
```

WAIS Uses: Documentation

- Up-to-date documentation
- Online help system
- Distribution of bug notices and fixes
- Mailing list archives

```
CMNS.Think.com  CM-Fortran.src
Quake.think.com wais-talk.src
PRISM CM programming environment
```

Conclusion

- Electronic Publishing can fill niches now
- Companies are positioning themselves now (workstations, server, and info providers)
- Thinking Machines is the "Engine of the Information Industry"

Conclusion

- Electronic Publishing can fill niches now
- Companies are positioning themselves now (workstations, server, and info providers)
- Thinking Machines is the "Engine of the Information Industry"